

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 17:32:52 ; Search time 1822.81 Seconds  
(without alignments)  
12358.072 Million cell updates/sec

Title: us-09-591-466c-1

Perfect score: 1669

Sequence: 1 gaattggcgccgcctgaga.....actttggcgccgcgaattc 1669

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.\*

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	719.8	43.1	768	10	BG594788	BG594788 EST493466
2	694	41.6	724	10	BI179808	BI179808 EST520753
3	638.6	38.3	661	10	BG889872	BG889872 EST515723
4	631.4	37.8	642	10	BG592816	BG592816 EST491494
5	615.8	36.9	643	9	A1486771	A1486771 EST245093
6	598.6	35.9	673	9	AW979500	AW979500 EST310521
7	586.2	35.1	659	9	AW735827	AW735827 EST336595
8	549	32.9	574	9	AW931180	AW931180 EST357023
9	479.6	28.7	602	10	BG131020	BG131020 EST463912
10	441.6	26.5	575	10	BG887316	BG887316 EST513167
11	431.6	25.9	681	9	A1725875	A1725875 BNLGH1133
12	423.8	25.4	528	9	A1897249	A1897249 EST266608
13	408	24.4	488	10	BG888733	BG888733 EST514584
14	395	23.7	530	9	AW618699	AW618699 EST320685
15	392.4	23.5	629	10	BE426174	BE426174 WHB0329_H
16	384.6	23.0	507	9	A1773401	A1773401 EST256501
17	380.4	22.8	664	10	BE805530	BE805530 ss47a01.Y

18	363.4	21.8	591	10	BG409413	BG409413 00785 lea
19	327.2	19.6	603	10	BE516817	BE516817 WHE620_D0
20	312.6	18.7	506	10	BG726055	BG726055 sae06c11.
21	288.6	17.3	488	10	BG508571	BG508571 sac74e07.
22	284	17.0	525	10	BM177681	BM177681 saj64c05.
23	277.4	16.6	418	10	BF066219	BF066219 st09b06.Y
24	261.4	15.7	472	10	BG051233	BG051233 FM1_57_H0
25	259.6	15.6	701	10	BJ170544	BJ170544 BJ170544
26	256.2	15.4	635	10	BF145860	BF145860 WHE18333_D
27	238.6	14.3	691	10	BE821775	BE821775 GW700015A
28	222.6	13.3	588	10	BE404993	BE404993 WHE1208_F
29	216.4	13.0	495	10	BG154089	BG154089 409_L1N01
30	212	12.7	388	10	BE325650	BE325650 NF089F02S
31	212	12.7	538	10	BE920532	BE920532 EST424301
32	207.8	12.5	227	10	BE461505	BE461505 EST412924
33	205.6	12.3	460	10	BG043733	BG043733 sv27e03.Y
34	200.8	12.0	2681	11	AK004760	AK004760 Mus muscu
35	189.6	11.4	448	10	BG159494	BG159494 OV2_5_A09
36	188.4	11.3	464	9	AW695874	AW695874 NF099F01S
37	174	10.4	386	10	BG046010	BG046010 saa47e07.
38	167.4	10.0	627	10	BM359759	BM359759 GA_Ea002
39	167.4	10.0	894	10	BG680531	BG680531 602628355
40	166.2	10.0	629	10	BG444140	BG444140 GA_Ea002
41	163.2	9.8	801	10	BI152075	BI152075 602916527
42	162.6	9.7	546	9	AW670629	AW670629 114751 MA
43	161	9.6	889	10	BI415357	BI415357 602988551
44	158.6	9.5	756	10	C96708	C96708 C96708 Rice
45	156.6	9.4	677	10	BE289472	BE289472 601087930

## ALIGNMENTS

RESULT	1	768 bp	mRNA	linear	EST 12-APR-2001
BG594788	EST493466	cSTS Solanum tuberosum	cdna clone	CSTS8020	5' sequence,
LOCUS	BG594788	768 bp	mRNA	linear	EST 12-APR-2001
DEFINITION	BG594788	768 bp	mRNA	linear	EST 12-APR-2001
ACCESSION	BG594788	768 bp	mRNA	linear	EST 12-APR-2001
VERSION	BG594788.1	GI:13612928			
KEYWORDS	EST.				
SOURCE	potato.				
ORGANISM	Solanum tuberosum				
REFERENCE	1 (bases 1 to 768)				
AUTHORS	van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemingo, A., Bougri, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.				
TITLE	Generations of ESTs from sprouting potato eyes				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Cathy Ronning The Institute for Genomic Research For clone info: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com Seq primer: M13F-R.				
FEATURES	Location/Qualifiers				
Source	1..768				
	/organism="Solanum tuberosum"				
	/cultivar="Kennebec"				
	/db_xref="taxon:4113"				
	/clone="CSTS8020"				
	/clone.lib="cSTS"				
	/tissue_type="sprouting eyes from tubers"				
	/dev_stage="12-14 weeks post harvest"				
	/lab_host="SOLR"				
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Various sizes of sprouting eyes (2mm to 15mm) were taken from tubers. The tubers were incubated at 26C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers."				
BASE COUNT	220 a 149 c 181 g 218 t				

## ORIGIN

Query Match	43.1%	Score 719.8;	DB 10;	Length 768;
Best Local Similarity	99.0%;	Pred. No. 2.4e-186;		
Matches 724;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
QY	15	ctgagaaacctgaattcaatttcgcatttgcagagatgagaggaaacaagtlltgc	74	
Db	36	CTGAGAAACCTCGAATTCAATTTCCGATTTGGCAGAGATGAGAGGAACAAGTTTGTCT	95	
QY	75	ttgatttacggtacaccttcctgcgtcggtgcgtctgccttcattcatcatcacagatcgggc	134	
Db	96	TTGATTTACGGTACCTTCTCGTCGTGCTCTCGCTTCATCTACATACAGATGGCG	155	
QY	135	ttttcgagcacagtcagaatatatgacacgccttgctgtgcatttgaagcagaaatc	194	
Db	156	TTTTTCGGGCACACAGTCAGAATATGTAGACGCCCTTGCTGTCAATTTGAAGCAAAATC	215	
QY	195	attgtacaagtcagaccagatgcttattcaacaagattagccagcagacaaggaagtag	254	
Db	216	ATTGTACAGTCAGCCACAGATTGCTTATTGACAAGATTAGCCACGACGAGGAAGATG	275	
QY	255	tagctcttgaagacaataatgaagcatcagaccaggagtgccggcaattaaaggctcttg	314	
Db	276	TAGCTCTTGAAGACAATAATGAAGCTCAGGACAGGAGTGCCGGCAATTAAGGCTCTTG	335	
QY	315	tcagatcttgaagtaagggcataaaaaagttaatcggagatgtcagatgccagtgg	374	
Db	336	TTCAGGATCTTGAAGTAAAGGATGAAAGAAAGTTAATCGGAGATGTGCAGATGCCAGTGG	395	
QY	375	cagctgtagttgttatggcttgagtcgtactgactacctggagaggactataaatcca	434	
Db	396	CAGCTGTAGTTGTTATGGCTTGAGTGTGCTGCTGACTACCTGGAGAGGACTATTAATTC	455	
QY	435	tcftaaataaccaaacatctgtgtgcatacaaatatcctctttcatatccagatggat	494	
Db	456	TCFTAAATAACCAACATCTGTTGCATCAAAATATCCTCTTTTCATATCCAGGATGGAT	515	
QY	495	caaatcctgatgtcaagaaagcttgccttttgactatgctcagctgaagtatatgcagact	554	
Db	516	CAATCTGATGTAAAGAACTTGCTTTTGAGCTATGATGACTGAGCTATATGCAGCACT	575	
QY	555	tggattatgaacctgtgcatactgaaagacagggaaactgggtgcatactacaagattg	614	
Db	576	TGGATTTTGAACCTGTGCATCTGAAAGACACAGGGAACTGGTTGCATACTACAAGATTG	635	
QY	615	cagctcattacaagtgggcatgtgattcagctgtttcaagaacataattttgacgctgta	674	
Db	636	CAGCTCATTTACAAGTGGGATTTGGATCAGCTGTTTTCACAAGCATAAATTTAGCCGTGTTA	695	
QY	675	tcatactagaagatgataggaaattgctgctgatttttttgaactattttgaagctggag	734	
Db	696	TCATACTAGAGATGATATGGAANATGCTGCTGATTTTTTTTGACTATTTTGGAGGCTGAG	755	
QY	735	ctactcttctt	745	
Db	756	CTACTCTCTCT	766	

## RESULT

LOCUS	DEFINITION	724 bp	mRNA	linear	EST 09-JUL-2001
BI179808	EST520753 cSTE Solanum tuberosum mRNA sequence.				
BI179808	BI179808				
ACCESSION	BI179808.1	GI:14645619			
VERSION	EST.				
KEYWORDS					
SOURCE	potato.				
ORGANISM	Solanum tuberosum				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanales; Solanales; Solanaceae; Solanum.				

REFERENCE  
AUTHORS

Chiemingo, A., Bougri, O., Buell, C. R., Ronning, C., Tanksley, S. and Baker, B.

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Research	1980, Vol. 83, No. 2, pp. 11-20
3. The Importance of Parental Involvement	Journal of Educational Research	1980, Vol. 83, No. 3, pp. 21-30
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 4, pp. 31-40
5. The Role of the School in the Community	Journal of Educational Research	1980, Vol. 83, No. 5, pp. 41-50
6. The Impact of Teacher Education on Classroom Practice	Journal of Educational Research	1980, Vol. 83, No. 6, pp. 51-60
7. The Importance of Teacher Collaboration	Journal of Educational Research	1980, Vol. 83, No. 7, pp. 61-70
8. The Effect of Teacher Attitudes on Student Behavior	Journal of Educational Research	1980, Vol. 83, No. 8, pp. 71-80
9. The Role of the School in the Development of the Child	Journal of Educational Research	1980, Vol. 83, No. 9, pp. 81-90
10. The Impact of Teacher Training on Student Learning	Journal of Educational Research	1980, Vol. 83, No. 10, pp. 91-100

The Institute for Genomic Research  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email [cdna@resgen.com](mailto:cdna@resgen.com)  
Seq primer: M13F-R.

FEATURES	
SOURCE	

```

/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="CSTE20K20"
/clone_lib="CSTE"
/tissue_type="axillary buds of stem explants; growing
sink-tubers"
/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"

```

/note="Vector: pBluescript SK(+); Site\_1: EcoRI; Site\_2: XhoI. Tissue supplied by Christian Bachem and Richard Visser (Department of Plant Breeding, Wageningen University, The Netherlands). The cSTA libraries will attempt to capture the induction and initiation/initial growth of the tuber in an *in vitro* system as described in Bachem et al. (1996). Small microtubers develop from axillary buds attached to stem explants when placed on a high sucrose medium (10%). Visible morphological changes occur synchronously at day five in the axillary buds. The first library, cSTA (1-20) consists of axillary buds harvested on days 1-3. This targets those genes involved in induction of the microtubers. The following libraries, cSTA (21-40) and cSTA (41-60), capture genes involved in tuber initiation and outgrowth. This library is noted as p3 in Tankslev lab notebooks".

BASE COUNT  
ORIGIN

```
Query Match          41.6%; Score 694; DB 10; Length 724;
Best Local Similarity 100.0%; Pred. No. 2.9e-179;
Matches 694; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

[illegible]

Qy	435	tcttaaaataccaaacatctgttcacataaaataatccctcttttccatccccagatgat	494
Db	451	TCCTAAAAATACCAACATCTCTTGCATCAAAATATCCTCTTTTCATATCCAGGATGGAT	510
Qy	495	caaatcctgatgaagaagctctcttgaactgatgtcagctaacgtatatcgacgact	554
Db	511	CAATATCCCTGATGTGAAGAAAGCTGCTTTGAGCTATGTGGTCAAGCTGACGTATATGACGACCT	570
Qy	555	tggattatgaacctgtgcatactgaagacacaggggaactggttgcatcactacaagattg	614
Db	571	TGGATTATGAACCTGTGCATACCTAAAGACACAGGGGAACCTGGTTGCATCTACAAGATTG	630
Qy	615	cagctcattacaagtgggcatttgatcagctgcttttcacaagcataaatttagccgctgta	674
Db	631	CACGTCTATCAAGTGGGCAATGGATCAGCTGCTTTCACAAGCATAAATTTAGCCCTGTTA	690
Qy	675	tcatactagaagatgatggaattgctgctga	708
Db	691	TCATATAGAGATGATATGGAATTCGTCGTA	724
RESULT 3			
BG889872			
LOCUS	EST515723	cSTD Solanum tuberosum cDNA clone cSTD15H10 5' sequence,	661 bp mRNA linear EST 30-MAY-2001
DEFINITION	BG889872	mRNA sequence.	
ACCESSION	BG889872		
VERSION	BG889872.1	GI:14266958	
KEYWORDS			
SOURCE		potato.	
ORGANISM		Solanum tuberosum	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.	
AUTHORS		van der Hoeven,R.; Bezzerides,J., Ewing,E., Cho,J., Chiémengo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.	
TITLE		Generations of ESTs from dormant potato tubers	
JOURNAL		Unpublished (2001)	
COMMENT		Contact: Cathy Ronning The Institute for Genomic Research For clone info: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com Seq primer: M13P-R.	
FEATURES		Location/Qualifiers	
source		1..661	
		/organism="Solanum tuberosum"	
		/cultivar="Kennebec"	
		/db_xref="taxon:4113"	
		/clone="cSTD15H10"	
		/clone_lib="cSTD"	
		/tissue_type="dormant tuber"	
		/dev_stage="one month post-harvest"	
		/lab_host="SOLR"	
		/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2: XhoI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tubers, avoiding the buds and epidermis. Tubers were stored for one month post-harvest at 40C. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as P4 in Tanksley lab notebooks."	
BASE COUNT	196 a	133 c	158 g 174 t
ORIGIN			
Query Match		38.3%;	Score 638.6; DB 10; Length 661;
Best Local		Similarity 99.4%;	Pred No. 4.4e-164;
Matches	641; Conservative	0; Mismatches	4; Indels 0; Gaps 0;
Qy	15	ctgagaacctgaatttcgatttcgagatgagaggaaacagtttgcgt	74

/tissue\_type="sprouting eyes from tubers"  
/dev\_stage="12-14 weeks post harvest"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were  
taken from tubers. The tubers were incubated at 26C in the  
dark for 2-3 weeks prior to sprouting. The eyes were  
frozen in liquid nitrogen immediately upon removal from  
tubers."  
BASE COUNT 180 a 124 c 145 g 193 t  
ORIGIN

Query Match 37.8%; Score 631.4; DB 10; Length 642;  
Best Local Similarity 99.1%; Pred. No. 4.1e-162;  
Matches 635; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 346 gtaatcgagagatgcagatgccagtcgagctgtagtgtattgcttgacgtcgatc 405  
Db 2 GTTAATCGGAGATGTCAGATGCCAGTGGCAGCTGTAGTTATGGCTTCGATCGTGC 61  
QY 406 tgactactggagaggactataaaccatcttaaaataccaaacatctgttgatcaaa 465  
Db 62 TGACTACTGGAGAGGACTATTAAATCCATCTTAAATACCAACATCTGTGCATCAA 121  
QY 466 atattcttttccatccaggatgagatcaaatccctgatgtaagaagcttgcttgag 525  
Db 122 ATATCTCTTTTCATATCCAGGATGGATCAAAATCCCTGATGTAAGAAAGCTTGTCTGAG 181  
QY 526 ctatggctcagctgacgtatgcagcacttgattgattgaacctgtgatacctgaagacc 585  
Db 182 CTATGATCAGCTGACGTATATGCAGCACTTGGATTTCACCTGTGCATCTGAAGACC 241  
QY 586 aggggaactggttgactactacagatgcacgtcatcaaatgagggcattggatcagct 645  
Db 242 AGGGAACTGGTTGCATACATACAGATTCGACGTCATTACAAAGTGGCAATGGATCAGCT 301  
QY 646 gttccacaagcataattttagccgtgttatcactactagaagatgatatgaaattgctgc 705  
Db 302 GTTTCACAAGCATAAATTTAGCCGTGTATCATACTAGAAGATGATATGGAATTTGCTGC 361  
QY 706 tgattttttgactattttgagctgagctactcttcttgacagagacaagtcgattat 765  
Db 362 TGATTTTTCGACTATTTTCAGGCTGGAGCTACTCTTCTTGACAGACACAAGTCGATTAT 421  
QY 766 ggtattttcttcttgaatgacaatggacaaaggcagctgctgcacagatcctgatgctct 825  
Db 422 GGCTATTTCCTTGGAAATGACAAATGGACAAAGSCAGTTCCTCCACAGATCCTGATGCTCT 481  
QY 826 ttaccgctcagacttttttctggtcttgatgagcttcttcaaaatcaacttggtccga 885  
Db 482 TTACCGCTCAGACTTTTTCCTGGCTTGGATGGATGCTTTCAAAATCAACTTGGTCCGA 541  
QY 886 actatctccaaagtggccaaaggcttactgggagctgagctgaagctgaaagaaatca 945  
Db 542 ACTATCTTCAAAGTGGCCAAAGCTTACTGGGATGACTGGCTGAAGCTGAAGAAATCA 601  
QY 946 cagagatcaacaatttatttcgccagaagtttcagaacgt 986  
Db 602 CAGAGGTGCAAAATTTATTACCCAGAAAGTTTGCAAAACGT 642

RESULT 5  
AI486771  
LOCUS  
DEFINITION EST245093 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
AI486771  
ACCESSION cLED11D20, mRNA sequence.  
AI486771  
VERSION AI486771.1 GI:4382142  
KEYWORDS EST  
SOURCE Tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon  
1 (bases 1 to 643)  
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,  
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman  
C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley  
S.D. and Giovannoni,J.  
Generation of ESTs from tomato carpel tissue  
Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.  
Location/Qualifiers  
source 1..643  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLED11D20"  
/clone\_lib="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="XL1-Blue MRF"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and  
directionally cloned cDNA in vector Lambda ZAP II with 5'  
and 3' ends located at the EcoRI and XhoI sites,  
respectively."  
BASE COUNT 179 a 128 c 144 g 192 t  
ORIGIN

Query Match 36.9%; Score 615.8; DB 9; Length 643;  
Best Local Similarity 97.4%; Pred. No. 7.8e-158;  
Matches 626; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 346 gtaatcgagagatgcagatgccagtcgagctgtagtgtattgcttgacgtcgatc 405  
Db 1 GTTAATCGGAAATGTGCAGATGCCAGTGGCAGCTGTAGTTTATGGCTTCGATCGTTC 60  
QY 406 tgactactggagaggactataaaccatcttaaaataccaaacatctgttgatcaaa 465  
Db 61 TGACTACTGGAGAGACTATAAAATCCATCTTAAATACCAACATCTGTTGCATCAA 120  
QY 466 atattcttttccatccaggatgagatcaaatccctgatgtaagaagcttgcttgag 525  
Db 121 ATATCTCTTTTCATATCCAGGATGGATCGAATCCCTGATGTTAGAAAGCTTGTCTTGAG 180  
QY 526 ctatggctcagctgacgtatgcagcacttgattgaacctgtgcatactgaaagacc 585  
Db 181 CTATGATCAACTGACGTATATGCAGCACTTGGATTTTGAACCTGTGCATCTGAAGACC 240  
QY 586 aggggaactggttgactactacagatgcacgtcatcaaatggcagcttgatcagct 645  
Db 241 AGGGAACTGGTTCATCTACTACAAGATTGCACGCTATTACAAGTGGCATTCGATCAGCT 300  
QY 646 gttccacaagcataaattttagccgtgttatcactactagaagatgatatgaaattgctgc 705  
Db 301 GTTTCACAAGCATAAATTTTAGCCGTGTATCATACTAGAAGATGATATGGAATTCGGCC 360  
QY 706 tgattttttgactattttgagctgagctactcttcttgacagagacaagtcgattat 765  
Db 361 TGATTTTTCGACTACTTTTCAGGCTGGAGCTACTCTTCTTGACAGACACAAGTCGATTAT 420  
QY 766 ggtattttcttcttgaatgacaatggacaaaggcagcttgcgtccaaagatcctgatgctct 825  
Db 421 GGCTATTTCATCTTGGAAATGACAAATGGACAAAGSCAGTTCCTCCACAGATCCTTATGCTCT 480  
QY 826 ttaccgctcagacttttttctggtcttgatgagcttcttcaaaatcaacttggtccga 885  
Db 481 TTACCGCTCAGACTTTTTCCTGGCTTGGATGGATGCTTTCAAAATCAACTTGGTCCGA 540



Qy 886 actatctccaaagtgccaaaggcttactgagtgactgctgaagctcaagagcaaatca 945  
|||||  
Db 541 ACTATCTCAAAGTGGCGAAGCCTTACTGGGATGACTGGCTGAGGCTCAAGCAAAATCA 600  
Qy 946 caagagtcagaattattcgcgccagaagtttgcgagaactac 988  
|||||  
Db 601 CAGAGGTCACAAATTTATTGCGCCAGAGAGTTTGCAGAACGTTTC 643

RESULT 6  
AW979500 673 bp mRNA linear EST 18-MAY-2001  
LOCUS EST310521 tomato root deficiency, Cornell University Lycopersicon  
DEFINITION esculentum cDNA clone cLEW209 5', mRNA sequence.  
ACCESSION AW979500  
VERSION AW979500.1 GI:8171019  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 673)  
van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,F.,  
Upton,J., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,  
Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.  
and Tanksley,S.D.  
Generation of ESTs from tomato nutrient-deficient roots  
Unpublished (1999)  
Contact: CUGI  
Clemson University  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
Location/Qualifiers  
1 . 673  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEW209"  
/clone\_lib="tomato root deficiency, Cornell University"  
/tissue\_type="roots"  
/dev\_stage="5-6 weeks old"  
/note="vector: pBluescript SK-; Site\_1: 5' EcoRI; Site\_2:  
3' XhoI; supplier: Tanksley; Tissue supplied by Dave  
Garvin (USDA-ARS, Ithaca, NY 14850). Roots were harvested  
from plants grown under the following  
deficiencies/stresses: 10 mM Al, Zn, P, K, Fe, N), and  
mRNA was isolated from individual treatments. Proportional  
aliquots of mRNA of each treatment were mixed and used for  
library construction."

BASE COUNT 198 a 136 c 159 g 180 t  
ORIGIN

Query Match 35.9%; Score 598.6; DB 9; Length 673;  
Best Local Similarity 96.2%; Pred. No. 4.2e-153;  
Matches 613; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 20 aaaccctcgaattcaatttcgcatttggcagagatgagagggaacaagtcttcttgat 79  
|||  
Db 37 AACCCTTAGAAATTCACATTTGCCATTTGGCAGAGATGAGAGCAACAAAGTTTGTCTTGAT 96  
Qy 80 ttacgggtaccttctcgtcggtgctctcgcttctatctacatcacagatcgcgcttttc 139  
|||||  
Db 97 TTACGGTACCTTCTCGTGGTGGCTGCTCTCGCCTTCATCTACATACAGATGCGGCTTTC 156  
Qy 140 gcacacagtcagaatatgtagaccgcgcttgcgtcgaattgaagcagaaaaatcattgt 199  
|||||  
Db 157 GTGACACAGTCAGAAATATGACAGCCGCTTGCTGCTCGATGAGGAGAGAAATCATGT 216

Qy 200 acaagtcagaccagattgcttattgacaagattagccagcagcaaggaagtagtagct 259  
|||||  
Db 217 ACAAGTCAGACTAGGTGGCTTATTGACAAGATTAGCCAGCAGCAAGAGTAGTAGTGCCT 276  
Qy 260 cttgaagaacaaatgaagcatcagaccagagtgccgccaattaaagggtctcttcttcag 319  
|||||  
Db 277 CTTGAAGAACAATGAAGGCCAGGCCAGGAGTGGCGCAATTAAGGGCTCTTTGTTTCAG 336  
Qy 320 gatcttgaagtaaggggcataaaaaagtttaatcgagagatgtagcagatgccagtgccagct 379  
|||||  
Db 337 GATCTTTGAAAGTAAGGCATATAAAAAAGTTAATCGGAAATGTGCAGATGCCAGTGGCAGCT 396  
Qy 380 gtagtggtatggcttgagctgactgactactactctgagaggactattaaatccatctta 439  
|||||  
Db 397 GTAGTTGTTATGGCTTGCAGTCGTTCTGACTACCTGAGGAAGACTATATAAATCCATCTTA 456  
Qy 440 aaataccaaacatctgtgcatacaaaatcatctcttttcatatccccagagatgaatcaat 499  
|||||  
Db 457 AAATACCAACATCTGTTGCATCAAAATATCCTCTTTTATATCCAGGATGGATCGAAT 516  
Qy 500 cctgatgtaagaaagctgcttcttgagctatggtcagctgacgtatgtagcagcacttgat 559  
Db 517 CCTGATGTTAGAAAGCTTGCCTTTGAGCTATGATCAACTGACGTATATGCAGCACTTTGGAT 576  
Qy 560 tatgaacctgtgcatactactaaagaccaggggaaactggtgcatctactacaagatgtcacgt 619  
|||||  
Db 577 TTTGAACCTGTGCATACTCAAAAGACCAGGGAACCTGGTTGCATACTACAAGATTGCACGT 636  
Qy 620 cattacaagtggcattggtgacgtctgttttcaaacg 656  
Db 637 CATTACAAGTGGCATTGGATCAGCTGGTTCAACAGC 673

RESULT 7  
AW735827 659 bp mRNA linear EST 18-MAY-2001  
LOCUS EST336595 tomato flower buds 0-3 mm, Cornell University  
DEFINITION Lycopersicon esculentum cDNA clone cTOA5111 5', mRNA sequence.  
ACCESSION AW735827  
VERSION AW735827.1 GI:7642666  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 659)  
van der Hoeven,R.S., Bezzeredes,J.L., Matern,A.L., Holt,I.E., Liang  
F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W.,  
Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.  
Generation of ESTs from tomato flower tissue, 0-3 mm buds  
Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
Location/Qualifiers  
1 . 659  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cTOA5111"  
/clone\_lib="tomato flower buds 0-3 mm, Cornell University"  
/tissue\_type="flower"  
/dev\_stage="0-3mm buds"  
/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Tanksley; Flower buds and flowers were  
taken from greenhouse plants (4-8 wks old, TA496). They  
were immediately frozen in liquid nitrogen and then





QY	1249	acagtttgccatcttttgaagaatgaaagatggtgtaccacggcgagcatataaaggat	1308
Db	62	ACAGTTTGCCATTTTGAAGATGAAGCATGCTGTACACGGCGAGCATATAAAGGAT	121
QY	1309	agtaatttcggtttcaaacatctagacgtgtgtctcttcttccctgattctcttcg	1368
Db	122	AGTAGTTTCCGGTTTCAAAACATCTAGACGCTGTGTCTTGTGCCCCCTGATCTCTCG	181
QY	1369	acaacttgagttgaagatacttagcgaagataatgattggagccctgagcaacaattaga	1428
Db	182	ACAACCTGGAGTTGAAGTACTTACCGAAGATATGATTTGGAGCCCTGAGCAACAATTAGA	241
QY	1429	cttatttggaagatacacatttgaagagctgacacgaaagatatgactaccagtagctac	1488
Db	242	CTTATTGGTAGATACATTTGAAAGAGCTGACACGAAAGTATGACTACCAGTAGCTAC	301
QY	1489	atgcaacattttaaattgaatgaaggaaacccactgctattgttggatggatgaatca	1548
Db	302	ATGCAACATTTTAAATTGTAATGAAGGAACCCACTGCTTATTTGTTGGAATGGATGAATCA	361
QY	1549	tcaccacatccctattattcaag-tttcaaacataaagaggaaatgttgcctataaaaa	1607
Db	362	TCACCACATCCATTATTCAAGTTTTCACAAACATAAAGAGGAAATGTAGCCCTATAAACA	421
QY	1608	caaatcttttcttaagaaggaacgttacgattatgagcaacttgg	1655
Db	422	CAATGTTTTTGTCTTAAAGGAACGTTAGCATTTTGAGCAACTTTG	469
RESULT	11		
LOCUS	AI725875	681 bp mRNA linear	EST 11-JUN-1999
DEFINITION	BNLGH113374 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to ALPHA-1, 3-MANNOSYL-GLYCOPROTEIN, mRNA sequence.		
ACCESSION	AI725875		
VERSION	AI725875.1	GI:5044727	
KEYWORDS	EST		
SOURCE	upland cotton.		
ORGANISM	Gossypium hirsutum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		
AUTHORS	Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.		
TITLE	ESTs from developing cotton fiber		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Ben Burr Biology Department Brookhaven National Laboratory Upton, NY 11973, USA Tel: 516-344-3396 Fax: 516-344-3407 Email: burr@bnl.bnl.gov		
FEATURES	Seq primer: T3 Primer. Location/Qualifiers		
source	1. 681 /organism="Gossypium hirsutum" /cultivar="Acala Maxxa" /db_xref="taxon:3635" /clone_lib="Six-day Cotton fiber" /tissue_type="immature fiber" /dev_stage="Six days post anthesis" /lab_host="XLI-Blue" /note="Vector: pBluescript II KS+"		
BASE COUNT	202 a 129 c 151 g 197 t	2	others
ORIGIN			
Query Match	25.9%	Score 431.6;	DB 9; Length 681;
Best Local Similarity	77.1%	Pred. No. 2.8e-107;	
Matches 524; Conservative	0;	Mismatches 136;	Indels 0; Gaps 0;
QY	506	gtagaagaagcttgcttgagctatgctgcagctgacgtatgacgacttgattatgaa	565

Db	1	GTAAAACTAAGCGCTTGAGTTATAGGAGCTAACTTATATGACGACATAGATTATGAT	60
QY	566	ctgtgcatactgaaagaccagggaactggttgcatactacaagatlgcacgtctattac	625
Db	61	CCGTGTCATACAGACGGCGCTGGGAAATTGATCGCATACTACAAGATTGCCGCTACTAC	120
QY	626	aagtgggcatgtgagtcagctgtttcacaaagcataattttagccgtgtatcatactacagaa	685
Db	121	AAATGGCATTTGGATGAGTTGTTCTACAAGACCAATTTTGGACCGAGCTAATAATACATTGAA	180
QY	686	gatgatatggaaattgctgctgatttttttgactattttgaggtcggagctactctctt	745
Db	181	GATGATATGGAATTTGCCCTGATTTTGTGATTACTTTGAGCGAGCTCTGCCCTTCTC	240
QY	746	gacagagacaagtcgattatgctattctcttgaatgacaatggacaaaggcagtlc	805
Db	241	GACAAGGACAAGTCAATTTATGCTGTTTCTCATGGAATGACAATGGCGCAAAAGCAGTTT	300
QY	806	gtccaagatccgtgctctttaccgctcagacttttttccctgggtcttggatggatgctt	865
Db	301	GTGTATGACCCCATATGCACCTTTATCGCTCAGATTTCTTTCTGGTCTTGGTGGATGCTT	360
QY	866	tcaaaatacaacttggccgaactatctccaaagtggccaaaggcttacttgggagtaactgg	925
Db	361	ACTAAATCTGTATGGAATGAGCTATCACAAAATGGCCAAAAGCTTACTTGGATGACTGG	420
QY	926	ctaaagctgaaagaaacacagaggtgcacaattattcgcgccagaagtcttgcagaacg	985
Db	421	TTGAGATTAAAAAATAATCACAAATGGTCGACAATTCCTTCGCTGGAAGTATGCAGAAC	480
QY	986	tacaatttggtagcatggttctagtttggggcagttttttaaagcagtatctttagacca	1045
Db	481	TATAATTTTGGTGAGCATGGTTCAAGCATGGGCGAGTTTTTCGAAAAATACCTTGCACCT	540
QY	1046	attaagctaaatgatgtccagttgatgtgaagtcaatgagaccttaagttaaccttttggag	1105
Db	541	ATTAAGATGAATGACGTGNAGGTGGACTCGAAGTNCAGGATTTTGAGCTTACCTTAAC	600
QY	1106	gacaactatgaaacacttttggcagcttggtaaaaggctaaagccctccacgagct	1165
Db	601	GAAAAATATGCCCAATACTTTCGACAGACATTTTCAMAGCGGCAAAACCTGTCCCTGGAACA	660
QY	1166	gatgctgttttgaagcatt	1185
Db	661	GATTCTGCTTTATGGCATT	680
RESULT	12		
LOCUS	AI897249	528 bp mRNA linear	EST 18-MAY-2001
DEFINITION	EST266608 tomato ovary, TAMU lycopersicon esculentum cDNA clone CLED26F21, mRNA sequence.		
ACCESSION	AI897249		
VERSION	AI897249.1	GI:5603067	
KEYWORDS	EST.		
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.		
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman ,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley ,S.D. and Giovannoni,J.		
TITLE	Generation of ESTs from tomato carpel tissue		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: CUGI Clemson University Clemson University 100 Jordan Hall, Clemson, SC 29634, USA		

Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
source  
Location/Qualifiers  
1..528  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLED26F21"  
/clone\_lib="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."  
BASE COUNT 171 a 94 c 119 g 144 t  
ORIGIN

Query Match 25.4%; Score 423.8; DB 9; Length 528;  
Best Local Similarity 89.2%; Pred. No. 3.5e-105;  
Matches 470; Conservative 0; Mismatches 52; Indels 5; Gaps 1;  
QY 1092 gtacactttggagacaactatgtgaacacatttggcgacttggtaaaaggctaagc 1151  
Db 1 GTTACCTTTTGGAGGACAACTATGTAAACACATTTTGGAGACTTGGTAAAGGCTAAGC 60  
QY 1152 ccacccagcgagctgactgttttgaagcatttaacatagatggatgctgattc 1211  
Db 61 CCATCCATGGAGCTGCTGCTTTGAAGCATTTAACATAGATGGTGCGTATTC 120  
QY 1212 agtacagagaccaactagactttgaagatatcgctcgacagtgttggcatttttgaagaat 1271  
Db 121 AGTACAGAGCAACTAGACTTTTGAAGACATCGCAGCACACTTTTGGAAATTTTCAAGAAT 180  
QY 1272 ggaagatggtgtaccagcgacacataaaggatagtagtttccgggttcaaacat 1331  
Db 181 GGAAGGATGTTGTACCGGCGCAGCATATAAGGAATAGTAGTTTCCGGTTTCAAAACAT 240  
QY 1332 ctacagctgtgtctctgtttccctgattctcttcgacaacttggagttgaagatactt 1391  
Db 241 CTACAGCTGTCTTCTTATTCGCCCTGATTCCTTCAACAACTTTGGAGTTCAAGATACTT 300  
QY 1392 agcaagatatgattggagcctgagcaacaatttagacttatttggtagatacatttga 1451  
Db 301 AGCAAAAGATATGATTGGAGCGGTGAGCAACAAATTTAGACTTATTTTGGCAGGATACATTTGA 360  
QY 1452 aagagctgacagaaaagtatgactaccagtagctacatgcatgcaacattttaattgaatgg 1511  
Db 361 AAGAGCTGGCAGCAAAAGCTTGATTAGCAGTAGCTGCATACACATTTTAAATGTAATGG 420  
QY 1512 aagaaaccactgcttattgttgaattggaatgcatcatcaccac-----atcctattatt 1566  
Db 421 AAGAACTCTACTGCTACTGTGGAGTGGATGAAATGTAGCCCCCATATAACACAATGTT 480  
QY 1567 caagtttacaacataaagagaaatttgcctctataaaacaaatt 1613  
Db 481 TTTCTTACAAACAAGAGGAATGTAGCCCCCATATAACACAAT 527

RESULT 13  
BG888733  
LOCUS  
DEFINITION EST514584 cSTD Solanum tuberosum cDNA clone cSTD11G24 5' sequence,  
mRNA sequence.  
ACCESSION BG888733  
VERSION BG888733.1 GI:14265819  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 488)  
Bouvier, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.  
Generations of ESTs from dormant potato tubers  
Unpublished (2001)  
Contact: Cathy Ronning  
The Institute for Genomic Research  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email [cdna@resgen.com](mailto:cdna@resgen.com)  
Seq primer: M13F-R.

FEATURES  
source  
Location/Qualifiers  
1..488  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone="cSTD11G24"  
/clone\_lib="cSTD"  
/tissue\_type="dormant tuber"  
/dev\_stage="one month post-harvest"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tuber, avoiding the buds and epidermis. Tubers were stored for one month post-harvest at 40C. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as P4 in Tanksley lab notebooks."  
BASE COUNT 146 a 86 c 118 g 138 t  
ORIGIN

Query Match 24.4%; Score 408; DB 10; Length 488;  
Best Local Similarity 96.5%; Pred. No. 7.2e-101;  
Matches 471; Conservative 0; Mismatches 10; Indels 7; Gaps 5;  
QY 1078 gtcaatggacctaagtacatttcttggagacaactatgtgaacacttttggcga--cttg 1135  
Db 1 GTCATGGACCTAAGTTACCTTTTGGAGGACAACATATGTGAACACATTTGGCGAACTGG 60  
QY 1136 gttaaaaggctaagcccatccacggagctgctgttttgaagcattttaacatagat 1195  
Db 61 TTAATAAGGGCTAAGCGCATCCAGGAGCTGATGCTGTTTGAAGACATTTTAACATAGAT 120  
QY 1196 ggtgattgctgtattcagtagacagacacactagactttgaagatatcgtctgcagattt 1255  
Db 121 GGTGATGTCGTATTTCAGTACAGAGACCACTAGACTTTGAAGCTATGCGTCGACAGTTT 180  
QY 1256 ggcatttttgaagaatggaaggat--ggtgtaccacggcgacgacataaaaggatagtagt 1314  
Db 181 GGCATTTTGAAGAATGGAAGGATGGGTGTACCACGGGACGACATATAAGGGATAGTAGT 240  
QY 1315 ttccgggtttcaaacactctagacgtgtgtctcttgtttccctgattctcttcgacaact 1374  
Db 241 TTTCCGGTTTCAAAACATCTAGACGTGTGTTCTCTTTGGCCCTGATTCTCTTCGACAAC 300  
QY 1375 tggagttgaagatacttagcgaagatatgattggagcctgagcaacaatttagacttatt 1434  
Db 301 TGGAGTTGAAGATACTTAGCGAAGATATGATTGGAGGCTGAGCAACAATTTAGACTTATT 360  
QY 1435 tggtaggatacatttgaagagctgcacgaaagttagtaccagtagctaccatgacacaa 1494  
Db 361 TGTTAGGATACATTGGAAGAGGCTGCACGAAAGATATGACTACCACTAGCTACATGCAA 420  
QY 1495 ca-ttttaattgaatgaagg-aacccactgcttattgttg--aatggatgaatcatc 1550  
Db 421 CATTTTAAATGTTAAATGGAAGAAACCCACTGCTTATTGTTGGGAAATGGGATGAATCATC 480  
QY 1551 accacatc 1558  
Db 481 ACCACATC 488

RESULT	14
AW618699	
LOCUS	530 bp mRNA linear EST 18-MAY-2001
DEFINITION	L. pennellii trichome; Cornell University Lycopersicon pennellii cDNA clone cLPT14D3 5', mRNA sequence.
ACCESSION	AW618699
VERSION	AW618699.1 GI:7324945
KEYWORDS	EST.
SOURCE	Lycopersicon pennellii.
ORGANISM	Lycopersicon pennellii
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
REFERENCE	Lycopersicon.
AUTHORS	1. (bases 1 to 530) Alcala, J., Vrebalov, J., White, R., Matern, A. L., Lakey, J., Holt, I. E., Liang, F., Hansen, T. S., Upton, J., Ronning, C. M., Craven, M. B., Fujii, C. Y., Bowman, C. L., Nierman, W., Fraser, C. M., Venter, J. C., Martin, G. B., Tanksley, S. D. and Giovannoni, J.
TITLE	Generation of ESTs from wild tomato ( <i>Lycopersicon pennellii</i> ) trichomes
JOURNAL	unpublished (1999)
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: <a href="http://www.genome.clemson.edu/orders/index.html">http://www.genome.clemson.edu/orders/index.html</a>

```

FEATURES
source
    prime sequence.
    Location/Qualifiers
      1..530
        /organism="Lycopersicon pennellii"
        /db_xref="taxon:28526"
        /clone="cLPT14D3"
        /clone_lib="f. pennellii trichome, Cornell University"
        /tissue_type="trichome"
        /dev_stage="mixed stages"
        /lab_host="SOLR"
        /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
        XhoI; Leaves of various stages were shaken in liquid
        nitrogen, shearing off trichomes. This procedure yielded a
        mixture of cells highly enriched for trichomes, with minor
        contamination by other types of leaf cells."
      155 a 102 c 128 g 145 t
BASE COUNT
ORIGIN

```

Query Match	23.7%	Score 395;	DB 9;	Length 530;
Best Local Similarity	96.4%;	Pred. No. 2.8e-97;		
Matches 404;	Conservative 0;	Mismatches 15;	Indels 0;	Gaps 0;
Qy 20	aaacccgcgaattcaaatcttcgacatttgccagagatgagagggaacaaagtcttgctttgat	79		
Db				
112	AAACCTTAGAATTCAATTTGCGATTGGCAGAGATGAGAGCAACAAGTTTGGCTTTGAT	171		
Qy 80	ttacggctacctcttcgctcgctggtgctctcgctcttcacatcacatcacagatgcggctcttc	139		
Db				
172	TTAGCGTACCTTCTCGTCGCTGCTCTCGCTTCTACATACAGATACGGGCTTTTC	231		
Qy 140	gcgacacagtcagaatatgtagaccgccttgctgctgcaattgaagcagaaaaatcattgt	199		
Db				
232	GTGACACAGTCAGAATATGAGACCCGCTTGCTGCTCAATTGAAGCAGAAAATCATTGT	291		
Qy 200	acaagtcagaccagattgctctattgacaagattagccagcagcaagaagagtagtagct	259		
Db				
292	ACAACTCAGACTAGTTTGCTTTATTGACAGATTAGCCAGCAGCAAGAGAGTAGTGGCT	351		
Qy 260	cttgaaagcaaatgaagcatcaggaccaggagtgccggcaattgaaggactcttgctcag	319		
Db				
352	CTTGAAGAACAATGAAGCGTCAGGACCAAGGAGTCCCGACAAATTAGGGCTCTTGTTTCAG	411		
Qy 320	gatcttgaagtgaagggcataaaaaagtttaactcggagatgtgcagatgccagtgccagct	379		

DB	412	GATCTTGAAGTAAGGGCATAAAAAGTTAATCGAAATGTGCAGATGCCACTGTCAGCT	471
QY	380	gtagttgttatggcttcgactgcgtactgactacctcgaggagactattaaatccatcitt	438
DB	472	GTAGTTGTTATGGCTTGCGAGTCGTCTCCTACCTACCTCGGAGAAGACTATAAAAATCCATCTT	530
RESULT	15		
BE426174			
LOCUS		629 bp mRNA linear EST 24-JUL-2000	
DEFINITION		WHE0329_H01_0012S Wheat unstressed seedling shoot cDNA library	
ACCESSION		Triticum aestivum cDNA clone WHE0329_H01_001, mRNA sequence.	
VERSION		BE426174	
KEYWORDS		BE426174..1 GI:9424017	
SOURCE		Estr.	
ORGANISM		Bread wheat.	
REFERENCE		Triticum aestivum	
AUTHORS		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum. 1 (bases 1 to 629) Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han .P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C. The structure and function of the expressed portion of the wheat genomes Unpublished (2000)	
JOURNAL		Contact: Olin Anderson	
COMMENT		US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: oanderse@w.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: Strataqene SK primer.	

[illegible]

```

/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0329_H01_001"
/clone_lib="Wheat unstrsed seedling shoot cDNA library"
/tissue_type="Etiolated shoot"
/dev_stage="Five day old seedling"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized,
germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Shoots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared, a cDNA library was made, and the cDNA clones
were in vivo excised to give pBluescript phagemids in the
TJ Clouse lab (Choi, Clouse, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
a 115 c 147 g 172 t

```

	Query Match	23.5%	Score 392.4;	DB 10;	Length 629;
	Best Local Similarity	76.7%	Pred. No. 1.5e-96;		
	Matches 480;	Conservative 0;	Mismatches 146;	Indels 0;	Gaps 0;
QY	455 gttgcatacaaaatatcctcttccatgccagatggatcaatccctgatgtaagaag				514
Db	2 GTTGCTTCAAAGTTTCACACTATTTATATCAGCATGGACAATAATGAGAGAAGTAAAAAG				61
QY	515 ctigctttgagctatggtcagctgacgtatcatcgagcacttgattgaacctgtgcatt				574

Db	62	AAAGCTTTGAGTTTACACTCAAAATAACATTTATCAGCATGTGGATCTTGAGCCTGTGGCC	121
Qy	575	actgaagacacagggaactgggttgatactactacaagatggcagctcattacaagtgga	634
Db	122	ACTCAAAACACGAGAGAAAACGTTGCATATTACAAGATAGCTAACCACTATATAATGGGCC	181
Qy	635	ttgatcagctggtttcacagaacaaatttagccggttatcatcatactagaagatgatg	694
Db	182	TTGATGATGAGCTATTTCATTAAAGCATGATTTTCGCGAGTAATCATTTCTGGAAGATCACATG	241
Qy	695	gaattgctgctgattttttgactattttgagctggagctactctcttgacagaagac	754
Db	242	GAGATCGCCCCAGATTCTTTGACTACTTTGAGGCTGCGAGCGAAATTACTTCACACTGAC	301
Qy	755	aagtcgattatggctatbttctcttggaatgacaatggacaaagcagttcgctccaagat	814
Db	302	AAGTCGATAATGGCTGTTTCTTCTTGGAAATGCAATGGGCAAAACGAGTTGTTTATGAC	361
Qy	815	ccctgagctctttaccgcctcagacttttctcgtcttgatggatgacgtttcacaatca	874
Db	362	CCAAAGCTCTTTTACCCTTGGGATTTCTTTCTCGGGCTTTGGAATGATGCTTAACCAAGTCA	421
Qy	875	acttgctccgaactatctccaaagtgggccaaaggcttactgggatgactgctgaagctg	934
Db	422	ACATGGATGGAGCTGTACACAAAGTGGCCCAAGCTTATTCGGATCACTGGGTGAGGCTA	481
Qy	935	aaagaaaaatcacagagggtcgacaatttatccgccagaagtttgcagacgtacaaat	994
Db	482	AAGGAGGTACAGAGATCGGCAGATTATTCGGCCAGAAAGTATGCAGAAACATACAAC	541
Qy	995	ggtgagcatggtctcagtttggggcagttttttaaagcagttatcttgagccaatt	1054
Db	542	GGCGACATGGATCAAGCATGGCACAATCTTTGATCAACTACTTGAACCAATCAAA	601
Qy	1055	aatgatgtccaggttctgattggaagtc	1080
Db	602	AATGATGCTCATATTTCGAGGAATTC	627

Search completed: August 13, 2002, 19:34:14  
Job time: 7282 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 17:55:12 ; Search time 83.73 Seconds  
(without alignments)  
4896.242 Million cell updates/sec

Title: US-09-591-466C-1

Perfect score: 1669

Sequence: 1 gaattcgccgcgcctgaga.....actttggcgcgcgaattc 1669

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	37.4	2.2	2934	1	US-08-198-446B-16
C 2	37.4	2.2	2934	2	US-08-870-693-16
C 3	37.2	2.2	7218	1	US-08-232-463-14
4	35.2	2.1	2939	4	US-09-276-531-80
5	35	2.1	1058	4	US-08-858-207A-127
6	34	2.0	1716	3	US-08-656-034-9
7	34	2.0	2160	3	US-08-656-034-1
8	33.8	2.0	3360	1	US-07-712-833A-1
9	33.6	2.0	9472	1	US-08-325-547-9
C 10	32.8	2.0	472	4	US-09-269-617-2
11	32.8	2.0	2441	1	US-08-920-812-12
12	32.8	2.0	2441	1	US-08-920-827-12
13	32.8	2.0	2441	1	US-08-921-177-12
14	32.8	2.0	2441	1	US-08-362-577C-12
15	32.8	2.0	2441	2	US-08-920-828-12
16	32.6	2.0	1183	1	US-08-188-582-25
17	32.6	2.0	1183	1	US-08-646-715-25
C 18	32.6	2.0	2801	3	US-08-747-221B-30
19	32.6	2.0	2801	3	US-08-747-221B-32
C 20	32.6	2.0	2801	4	US-09-005-051-30
21	32.6	2.0	2801	4	US-09-005-051-32
C 22	32.6	2.0	2836	3	US-08-747-221B-24
23	32.6	2.0	2836	3	US-08-747-221B-26
C 24	32.6	2.0	2836	4	US-09-005-051-24
25	32.6	2.0	2836	4	US-09-005-051-26
26	32.2	1.9	2265	3	US-09-369-618-3
27	32.2	1.9	2265	3	US-09-369-617-3

C 28	31.8	1.9	3479	3	US-08-714-918-7	Sequence 7, Appli
C 29	31.8	1.9	3479	4	US-09-265-315-7	Sequence 7, Appli
C 30	31.8	1.9	3479	4	US-09-265-315-7	Sequence 7, Appli
C 31	31.8	1.9	3479	4	US-09-266-417-7	Sequence 7, Appli
C 32	31.8	1.9	10968	2	US-08-680-327-2	Sequence 2, Appli
C 33	31.8	1.9	10968	4	US-09-228-246-1	Sequence 1, Appli
C 34	31.6	1.9	3116	1	US-08-149-103-2	Sequence 2, Appli
C 35	31.6	1.9	3116	1	US-08-451-883-2	Sequence 2, Appli
C 36	31.4	1.9	624	4	US-09-397-992A-6	Sequence 6, Appli
C 37	31.4	1.9	1868	2	US-08-960-022-5	Sequence 5, Appli
C 38	31.4	1.9	80246	4	US-09-078-294-4	Sequence 4, Appli
C 39	31.4	1.9	80595	4	US-09-078-294-3	Sequence 3, Appli
C 40	31.2	1.9	398	3	US-08-714-918-5	Sequence 5, Appli
C 41	31.2	1.9	398	4	US-09-265-315-5	Sequence 5, Appli
C 42	31.2	1.9	398	4	US-09-265-315-5	Sequence 5, Appli
C 43	31.2	1.9	398	4	US-09-266-417-5	Sequence 5, Appli
C 44	31	1.9	624	4	US-09-397-992A-3	Sequence 3, Appli
C 45	30.8	1.8	578	4	US-09-020-956-104	Sequence 104, App

ALIGNMENTS

RESULT 1  
US-08-198-446B-16/c  
; Sequence 16, Application US/08198446B  
; Patent No. 5674996  
; GENERAL INFORMATION:  
; APPLICANT: Hartwell, Leland H.  
; APPLICANT: Weinert, Ted A.  
; APPLICANT: Plon, Sharon E.  
; APPLICANT: Groudine, Mark T.  
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
; STREET: 1420 Fifth Ave., Suite 2800  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101-2347  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/198,446B  
; FILING DATE: 18-FEB-1994  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sheiness, Diana K.  
; REGISTRATION NUMBER: 35,356  
; REFERENCE/DOCKET NUMBER: FHCRI7537  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-682-8100  
; TELEFAX: 206-224-0779  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2934 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; DESCRIPTION: yeast MEC2 cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 395..2724  
; US-08-198-446B-16



```
Query Match          2.2%; Score 37.4; DB 1; Length 2934;
Best Local Similarity 48.4%; Pred. No. 0.16;
Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 621 attacaagtgccgattggatcagctggtttcacagaacataattttagccgtgttatcatc 680
    || || || || || || || || || || || || || || || || || || || || || ||
Db 2020 ATATCGAATAGGCGGTGTCGGGTATTTTAAATCTTGAATTTTCCATCTTGGTCTTC 1961

QY 681 taagaatgataatgaaatgctgctgattttttgactattttgagcgtggagctactc 740
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1960 CTGATCCTGTTCTTGAAGTTGTTGCTCCATTTCTTAATAATTTCTGCGCACAAAGATTGTCATAAGGA 1841

QY 801 agtctccaagatcctgatcgtctttaccgctca 835
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1840 TATTGTGAAAAATCACCATATGATTGTGAGGCCA 1806

RESULT 2
US-08-870-693-16/c
; Sequence 16, Application US/08870693
; Patent No. 5866338
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.
; APPLICANT: Weinert, Ted A.
; APPLICANT: Plon, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,693
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,446
; FILING DATE: February 18, 1994
; APPLICATION NUMBER: PCT/US93/04458
; FILING DATE: May 12, 1993
; APPLICATION NUMBER: US 07/884,426
; FILING DATE: May 14, 1992
; APPLICATION NUMBER: US 07/882,051
; FILING DATE: May 12, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCRI10798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2934 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
```

```
; DESCRIPTION: yeast MEC2 cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 395..2724
US-08-870-693-16

Query Match          2.2%; Score 37.4; DB 2; Length 2934;
Best Local Similarity 48.4%; Pred. No. 0.16;
Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 621 attacaagtgccgattggatcagctggtttcacagaacataattttagccgtgttatcatc 680
    || || || || || || || || || || || || || || || || || || || || || ||
Db 2020 ATATCGAATAGGCGGTGTCGGGTATTTTAAATCTTGAATTTTCCATCTTGGTCTTC 1961

QY 681 tagaagatgataatgaaatgctgctgattttttgactattttgagcgtggagctactc 740
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1960 CTGATCCTGTTCTTGAAGTTGTTGCTCCATTTCTTAATAATTTCTGCGCACAAAGATTGTCATAAGGA 1841

QY 741 tctttgacagagacaagtcgattatggtatttcttcttggaaatgacaatgggacaagcc 800
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1900 GTATTGAGCATCGTCCATATTTTCTTAATAATTTCTGCGCACAAAGATTGTCATAAGGA 1806

QY 801 agtctccaagatcctgatcgtctttaccgctca 835
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1840 TATTGTGAAAAATCACCATATGATTGTGAGGCCA 1806

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptz9pt-Fls
US-08-232-463-14

Query Match
Best Local Similarity 2.2%; Score 37.2; DB 1; Length 7218;
Matches 12; Conservative 150; Mismatches 108; Indels 0; Gaps 0;

Qy 153 aatagttagaccgcttgcgtgcaattgaagcagacagaaatattgtacaagtcacacca 212
Db 1456 AAGAGATAGAAGATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1397

Qy 213 gattgctattgacaagattagccagcagcagcagcagcagcagcagcagcagcagc 272
Db 1396 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1337

Qy 273 tgaagcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 332
Db 1336 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1277

Qy 333 agggcataaaagtttaacggagatgtgcagatgccagctgagcagctgtattgtatgg 392
Db 1276 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1217

Qy 393 ctgcagctgactgactaccctggagagga 422
Db 1216 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1187

RESULT 4
US-09-276-531-80
; Sequence 80, Application US/09276531
; Patent No. 6183968
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Reddy, Roopa
; APPLICANT: Guegler, Karl J.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
; TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,531
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/079,677
; FILING DATE: March 27, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lyon E. Murry, Ph.D.
; REGISTRATION NUMBER: 42,918
; REFERENCE/DOCKET NUMBER: PA-0008 US

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2939 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNUT06
; CLONE: 1611508
US-09-276-531-80

Query Match
Best Local Similarity 2.1%; Score 35.2; DB 4; Length 2939;
Matches 100; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 1301 aaaggagtagtatttcocgggtttcaaacatctagacgtgcttccttcttccccctgat 1360
Db 465 AAGGAGATACACTGACTCCCTGGTTTATTTGCATTAATAATTTATGTTGAGGTTACCTCAACT 524

Qy 1361 tctcttcgacaacttggagttgaagatacttagcgaagatatgattggagcctgagcaac 1420
Db 525 TGTTTTAAAGATTTTGTGTTGTAATTTGTACTGTATATTTGAGTAACCTGTCAGGCTTT 584

Qy 1421 aatttagcttatttggtaggatacatcttgaagagagctgacacgaaagtatgactaccca 1480
Db 585 TATTTAAATTTGTTTAACATGTACCATGTACCATGTACATGTACTTACTTATTTCAATGCATCATG 644

Qy 1481 gtactacatgcaacattttaatgttaa 1508
Db 645 CTGTAAACAGGCTTTTCATTATAATAA 672

RESULT 5
US-08-858-207A-127
; Sequence 127, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478

```



Qy 1414 gagcaacaatttagacttatttggtaggatacatttgaagagctgacacgaaaaagtatg 1473

```
Query Match      2.0%; Score 33.6; DB 1; Length 9472;
Best Local Similarity 50.6%; Pred. No. 5.3;
Matches 81; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
```

QY 1502 atgttaatggaagaaacccactgcttatttqtggaatggatgaatcatcaccacatccta 1561

```
Db 6125 ATATAGATTTAAGAAACCCGACGAGATTATGTTATGTTGGTGGTTCAGATTCTA 6184
Qy 1562 ttattcaagtttaacataaagaggaatgttgccctataaaacaaatttttgttt 1621
Db 6185 TTATTGGCAGTTGCAATCTACATTATCAATAACTACTGCTTAGAAATATTATATAATT 6244
Qy 1622 ctaagaaggaacgttacgattatgagcaactttggcgagcc 1661
Db 6245 TTCATTGTCACACAATAAGAATGCGAAGAGGATTGCGAGCC 6284

RESULT 10
US-09-269-617-2/c
; Sequence 2, Application US/09269617
; Patent No. 6204253
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: FACTORS THAT INTERACT WITH ONCOPROTEINS
; NUMBER OF SEQUENCES: 14
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/269,617
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: B-lymphocyte
; IMMEDIATE SOURCE:
; LIBRARY: Human cDNA library
; CLONE: SZ10
; US-09-269-617-2

Query Match 2.0%; Score 32.8; DB 4; Length 472;
Best Local Similarity 59.0%; Pred. No. 1.4;
Matches 49; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1568 agtttacaacataaagaggaatgttgccctataaaacaaatttttgtttcttaaga 1627
Db 453 AAAATTNNAAAAAANNNGGAAAGTTTCTTNNAAAAAATTTTNNCAAAAANTG 394
Qy 1628 aggaacgttacgattatgagca 1650
Db 393 GCGAACCCCGAGGANTAAATCGGAA 371

RESULT 11
US-08-920-812-12
; Sequence 12, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
```

```
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecalis
; STRAIN: Clinical isolate S2-27
; US-920-812-12

Query Match 2.0%; Score 32.8; DB 1; Length 2441;
Best Local Similarity 52.1%; Pred. No. 4.1;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy 743 cttgacagagacaagtcgattatgctattcttcttggaatgacaaatggacaaagcgag 802
Db 1848 CTTTCCTAAGATAACTAGAAATTTTCTTACGCTCAGAAACCAAGCTCAATTATTG 1907
Qy 803 ttcgtccaagatcctgatgcctctttaccgcctcagaatttttcttggtcttgatgatg 862
Db 1908 TGATTACCCCTATAATCTTCTTTTATTTCGGCGACCTCTTTAATATGATTAATTGGAG 1967
Qy 863 ctttcaaaaatcaacttggtc 882
Db 1968 TTTTAAATTTGAAGCTGTC 1987

RESULT 12
US-08-920-827-12
; Sequence 12, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
```

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/920.827  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2441 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORGANISM: Enterococcus faecalis  
; STRAIN: Clinical Isolate S2-27  
; US-08-920-827-12

Query Match 2.0%; Score 32.8; DB 1; Length 2441;  
Best Local Similarity 52.1%; Pred. No. 4.1;  
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 743 ctgtacagagacagtcgattatgctattcttcttgaatgacaaagcgag 802  
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |  
Db 1848 CTTTCCTAAGATAACTAGAAATTTTCTTACGCTCAGAAAGCCAAAGCTCAATTATTG 1907  
QY 803 ttgtccaagatcgtgatgctctttaccgctcagactttttccctggcttgatgatg 862  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1908 TGATTACCCATAATCTCTCTTTATTTCGGGACCTCTTTAATAATGATTAATTGGAGG 1967  
QY 863 ctttcaaatcaactggctc 882  
||| ||| ||| ||| |||  
Db 1968 TTTTAAATTGAAAGCTGTC 1987

RESULT 13  
US-08-921-177-12  
; Sequence 12, Application US/08921177  
; Patent No. 5798211  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/921,177  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2441 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORGANISM: Enterococcus faecalis  
; STRAIN: Clinical Isolate S2-27  
; US-08-921-177-12

Query Match 2.0%; Score 32.8; DB 1; Length 2441;  
Best Local Similarity 52.1%; Pred. No. 4.1;  
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 743 ctgtacagagacagtcgattatgctattcttcttgaatgacaaagcgag 802  
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |  
Db 1848 CTTTCTTAAGATAACTAGAAATTTTCTTACGCTCAGAAAGCCAAAGCTCAATTATTG 1907  
QY 803 ttgtccaagatcgtgatgctctttaccgctcagactttttccctggcttgatgatg 862  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1908 TGATTACCCATAATCTCTCTTTATTTCGGGACCTCTTTAATAATGATTAATTGGAGG 1967  
QY 863 ctttcaaatcaactggctc 882  
||| ||| ||| ||| |||  
Db 1968 TTTTAAATTGAAAGCTGTC 1987

RESULT 14  
US-08-362-577C-12  
; Sequence 12, Application US/08362577C  
; Patent No. 5807673  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,577C  
; FILING DATE: 27-MAR-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien

Query Match 2.0%; Score 32.8; DB 2; Length 244;  
Best Local Similarity 52.1%; Pred. No. 4.1;  
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy	743	cttgcagagacaagtcgattatgactattctctctggaatgacattggaacagcag	802
Db	1848	CTTTCCTAGATAACTAGAAATTTTCTCTAGCTCAGAAGCCAAAGCTCAATTATG	1907
Qy	803	ttcgtccaaagatactcgtatgctctttaccgcctcagactttttctctggcttggatgcaty	862
Db	1908	TGATTACCCCTAAATCTCTCTTTATTGCGGACCTCTTTAATATGATTAAATTCAGG	1967
Qy	863	ctttcaaaatcaacttggct	882
Db	1968	TTTTTAAATTTGAAGCTGTC	1987

```

RESULT 15
US-08-920-828-12
? Sequence 12, Application US/08920828
? Patent NO. 5853998
? GENERAL INFORMATION:
? APPLICANT: Ohno, Tsuneya
? APPLICANT: Matsuhisa, Akio
? APPLICANT: Uehara, Hirotsugu
? APPLICANT: Eda, Soji
? TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
? NUMBER OF SEQUENCES: 25
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
? STREET: 6300 Sears Tower, 233 South Wacker Drive
? CITY: Chicago
? STATE: Illinois
? COUNTRY: United States of America
? ZIP: 60606-6402
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? FILING DATE: 29-AUG-1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? FILING DATE: 27-MAR-1995
? APPLICATION NUMBER: US 08/362,577
? ATTORNEY/AGENT INFORMATION:
? NAME: Bin-Laures, Li-Hsien
? REGISTRATION NUMBER: 33,547
? REFERENCE/DOCKET NUMBER: 19036/32420
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 312/474-6300
? TELEFAX: 312/474-0448

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 18:14:32 ; Search time 285.38 Seconds  
(without alignments)  
10041.104 Million cell updates/sec

Title: US-09-591-466c-1  
Perfect score: 1669  
Sequence: 1 gaattgcgcgcgcctgaga.....actttgcgcgcgcgaattc 1669

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : N\_Geneseq\_032802.\*
- 1: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
  - 2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
  - 3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
  - 4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
  - 5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
  - 6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
  - 7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
  - 8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
  - 9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
  - 10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
  - 11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
  - 12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
  - 13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
  - 14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
  - 15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
  - 16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
  - 17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
  - 18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
  - 19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
  - 20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
  - 21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
  - 22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
  - 23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
  - 24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	1669	100.0	1669	20	AAx78001	Potato GntI cDNA.
2	1255	75.2	1737	20	AAx78002	Tobacco GntI cDNA.
3	751	45.0	1639	21	AAC39540	Arabidopsis thaliana
4	728.6	43.7	1641	21	AAC50197	Arabidopsis thaliana
5	329.2	19.7	510	20	AAx78003	A. thaliana GntI-h
6	192.8	11.6	2557	15	AAQ62625	Rat N-acetyl-gluco
7	181.6	10.9	2485	13	AAQ25594	Rabbit Gnt I cDNA
8	165.6	9.9	1335	7	AAI69837	Human beta-1,2-N-a
9	164	9.8	3230	23	AAQ25595	Human Gnt I cDNA C

10	144.2	8.6	2250	22	AAF93806	Human cDNA encodin
11	131.4	7.9	2367	23	ABL22893	Drosophila melanog
12	121.2	7.3	355	21	AAC69652	Human acetylglucos
C 13	58.8	3.5	4721	23	ABL22692	Drosophila melanog
14	44	2.6	2078	22	AAH15975	Human cDNA sequenc
15	44	2.6	2353	22	AAI58317	Human polynucleoti
16	44	2.6	2372	21	AAF18300	Lung cancer associ
17	44	2.6	2711	22	AAI60103	Human polynucleoti
18	44	2.6	2725	21	AAI37048	Human polynucleoti
19	44	2.6	2725	22	AAI21483	Human cDNA sequenc
20	44	2.6	2725	22	AAI21483	Human cDNA sequenc
21	44	2.6	2725	22	AAI21483	Human cDNA sequenc
22	44	2.6	2725	22	AAI21483	Human cDNA sequenc
23	44	2.6	2731	21	AAI21483	Human cDNA sequenc
24	41.2	2.5	725	22	AAI21483	Human cDNA sequenc
C 25	40.2	2.4	4590	22	AAI21483	Human cDNA sequenc
26	39.8	2.4	300	21	AAI21483	Human cDNA sequenc
27	39	2.3	11260	22	AAI21483	Human cDNA sequenc
C 28	38.8	2.3	11473	20	AAI21483	Human cDNA sequenc
C 29	38.8	2.3	11473	20	AAI21483	Human cDNA sequenc
C 30	38.4	2.3	343	21	AAI21483	Human cDNA sequenc
31	38	2.3	910715	20	AAI21483	Human cDNA sequenc
C 32	37.8	2.3	14041	22	AAI21483	Human cDNA sequenc
C 33	37.4	2.2	2934	18	AAI21483	Human cDNA sequenc
C 34	37.4	2.2	2934	18	AAI21483	Human cDNA sequenc
C 35	37.4	2.2	3299	22	AAI21483	Human cDNA sequenc
C 36	37.4	2.2	4900	22	AAI21483	Human cDNA sequenc
C 37	37.4	2.2	12842	22	AAI21483	Human cDNA sequenc
38	37.2	2.2	14316	24	AAI21483	Human cDNA sequenc
39	37	2.2	2328	20	AAI21483	Human cDNA sequenc
C 40	36.6	2.2	2265	22	AAI21483	Human cDNA sequenc
C 41	36.6	2.2	2265	22	AAI21483	Human cDNA sequenc
C 42	36.4	2.2	5059	20	AAI21483	Human cDNA sequenc
C 43	36.2	2.2	16768	23	AAI21483	Human cDNA sequenc
C 44	36	2.2	400	22	AAI21483	Human cDNA sequenc
C 45	36	2.2	7885	20	AAI21483	Human cDNA sequenc

ALIGNMENTS

RESULT 1  
AAx78001  
ID AAX78001 standard; cDNA; 1669 BP.

XX AAX78001;  
AC AAX78001;

DT 19-AUG-1999 (first entry)  
XX

DE Potato GntI cDNA.  
XX

KW GntI: beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;  
KW deficient; defective; detection; transgenic plant; sugar residue;  
KW medicine; antigenicity; deglycosylation; potato; ds.  
XX Solanum tuberosum.  
XX

FT Key Location/Qualifiers  
CDS 53..1394  
FT /\*tag= a  
FT /product= "GntI"

XX DE19754622-A1.  
XX

PD 10-JUN-1999.  
XX

PF 09-DEC-1997; 97DE-1054622.  
XX

PR 09-DEC-1997; 97DE-1054622.  
XX

PA (VSCH/) VON SCHAEWEN A.  
XX

XX Von Schaeuwen A;  
PI



XX WPI; 1999-338905/29.  
DR P-PSDB; AAY0888.  
XX  
XX Nucleic acid encoding plant N-acetylglucosaminyl transferase I  
PT useful for generating plants producing glycosylated proteins of low  
PT antigenicity  
XX  
XX Claim 6; Page 16-18; 37pp; German.  
XX  
XX This invention describes novel plant N-acetylglucosaminyl transferase I  
CC (GnTI) proteins and their encoding nucleic acids. The nucleic acid  
CC sequences of the invention may be used for recombinant production of the  
CC encoded proteins, which are then used to raise antibodies (Ab) for  
CC identifying plants with deficient or defective production of the protein.  
CC They may also be used to detect such plants by hybridization and to  
CC isolate related sequences from other plants or to generate antisense or  
CC sense constructs for reducing/deleting GnTI protein activity in plants.  
CC These transgenic plants may be used to produce glycoproteins with  
CC minimal, uniform and defined sugar residues. Such glycoproteins are  
CC useful in medicine and research, e.g. human glucocerebrosidase for  
CC treating Gaucher's disease. Plants which are defective or deficient in  
CC production of the GnTI protein can be made to produce glycoproteins with  
CC minimal, uniform and defined sugar residues, of low antigenicity. Use of  
CC these plants eliminates the need for the difficult isolation and  
CC deglycosylation of native proteins or preparation in defective animal  
CC cells. This sequence represents potato (Solanum tuberosum) GnTI encoding  
CC cDNA.  
XX  
XX Sequence 1669 BP; 489 A; 312 C; 387 G; 481 T; 0 other;  
SQ

Query Match 100.0%; Score 1669; DB 20; Length 1669;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 1669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gaattcgcgcgcctgagaaacctgaattcaatttcgacttgcagagatgagagg 60  
Db 1 gaattcgcgcgcctgagaaacctgaattcaatttcgacttgcagagatgagagg 60

Qy 61 gaacaagtttgcctgattacaggtacacctcttcgctggtgctctcgcttcaatc 120  
Db 61 gaacaagtttgcctgattacaggtacacctcttcgctggtgctctcgcttcaatc 120

Qy 121 catcacatgcgcttttcgcacacagtcagaatatgtagaccgccttgcgtgcaat 180  
Db 121 catcacatgcgcttttcgcacacagtcagaatatgtagaccgccttgcgtgcaat 180

Qy 181 tgaagcagaaatcattgtacaagtcagaccagattgtcttattgacaagattagccagca 240  
Db 181 tgaagcagaaatcattgtacaagtcagaccagattgtcttattgacaagattagccagca 240

Qy 241 gcaagggaagtagtagctcttgaagacaataatgaagcatcagaccagagtgccggca 300  
Db 241 gcaagggaagtagtagctcttgaagacaataatgaagcatcagaccagagtgccggca 300

Qy 301 attagggctctgttcagatcttaaaagtaaggggcataaaaaagtaaatcgagagatgt 360  
Db 301 attagggctctgttcagatcttgaagtaaggggcataaaaaagtaaatcgagagatgt 360

Qy 361 gcagatccagtggaagctgtagttgttgcttgcgtcgtactgactacctggagag 420  
Db 361 gcagatccagtggaagctgtagttgttgcttgcgtcgtactgactacctggagag 420

Qy 421 gactattaaatccatcttaaaataccacaacatctgttgcatacaaaatctcttttcat 480  
Db 421 gactattaaatccatcttaaaataccacaacatctgttgcatacaaaatctcttttcat 480

Qy 481 atcccagatggatcaaatccctgatgtaagaaagcttgcctttagctatggtcagctgac 540  
Db 481 atcccagatggatcaaatccctgatgtaagaaagcttgcctttagctatggtcagctgac 540

Qy 541 gtatatgcagcacttggattgaacctgtgcatactgaaagaccaggggaactggtgc 600

Db 541 gtatatgcagcacttggattgaacctgtgcatactgaaagaccaggggaactggtgc 600  
Qy 601 atactacaagattgcacgtcattcaacagtggcgattggatcagctgtttcacaagcataa 660  
Db 601 atactacaagattgcacgtcattcaacagtggcgattggatcagctgtttcacaagcataa 660  
Qy 661 ttttagccgtgttatcactagaagatgatatggaattgcgtgattttttttagacta 720  
Db 661 ttttagccgtgttatcactagaagatgatatggaattgcgtgattttttttagacta 720  
Qy 721 ttttagcgtggagctactcttcttgacagagacaagtcgattatggctatttcttcttg 780  
Db 721 ttttagcgtggagctactcttcttgacagagacaagtcgattatggctatttcttcttg 780  
Qy 781 gaatgacaatgacaaagcagtcgtccaagatcctgatgctcttaccgctcagactt 840  
Db 781 gaatgacaatgacaaagcagtcgtccaagatcctgatgctcttaccgctcagactt 840  
Qy 841 ttttctcgtgcttgatggatgcttttcaaaatcaacttggctccgaactatctccaagt 900  
Db 841 ttttctcgtgcttgatggatgcttttcaaaatcaacttggctccgaactatctccaagt 900  
Qy 901 gccaaagcttactgggagtgactgctgaagcgtgaaagaaatacacagaggtcgacaatt 960  
Db 901 gccaaagcttactgggagtgactgctgaagcgtgaaagaaatacacagaggtcgacaatt 960  
Qy 961 taticgccagagaagtttcagaaacatacaattttggtgagcatgcttctagtttgggca 1020  
Db 961 taticgccagagaagtttcagaaacatacaattttggtgagcatgcttctagtttgggca 1020  
Qy 1021 gtttttaagcagtatcttggagccaattaaagctaaatgatgccagggtgattggaagtc 1080  
Db 1021 gtttttaagcagtatcttggagccaattaaagctaaatgatgccagggtgattggaagtc 1080  
Qy 1081 aatggacctaaagttacotttttggaggacaactatgtgaaacactttggcgacttgggttaa 1140  
Db 1081 aatggacctaaagttacotttttggaggacaactatgtgaaacactttggcgacttgggttaa 1140  
Qy 1141 aaaggctaaagcccatccacggagctgctgctgttttgaagcattttaaactagatgggtga 1200  
Db 1141 aaaggctaaagcccatccacggagctgctgctgttttgaagcattttaaactagatgggtga 1200  
Qy 1201 tgtgcgtattcagtcacagagaccaactagacttgaagatatcgtcgcagagtttggcat 1260  
Db 1201 tgtgcgtattcagtcacagagaccaactagacttgaagatatcgtcgcagagtttggcat 1260  
Qy 1261 ttttgaagaaatggaagatggttaccacggcgagcatataaaggagatagttttccg 1320  
Db 1261 ttttgaagaaatggaagatggttaccacggcgagcatataaaggagatagttttccg 1320  
Qy 1321 gtttcaaacatctagacgctgtgttcccttgtttccctcgtattctctcgacaacttggagt 1380  
Db 1321 gtttcaaacatctagacgctgtgttcccttgtttccctcgtattctctcgacaacttggagt 1380  
Qy 1381 tgaagatctacttagcgaagatagattggagcctgagcaacaatttagacttatttggtag 1440  
Db 1381 tgaagatctacttagcgaagatagattggagcctgagcaacaatttagacttatttggtag 1440  
Qy 1441 gatacatttgaagagcgtgacacgaaaagatgactaccagtagctacatgcaacatttt 1500  
Db 1441 gatacatttgaagagcgtgacacgaaaagatgactaccagtagctacatgcaacatttt 1500  
Qy 1501 aatgttaatggaaggaacccactcttattgttggaaatggatgaatcatcaccacatcct 1560  
Db 1501 aatgttaatggaaggaacccactcttattgttggaaatggatgaatcatcaccacatcct 1560  
Qy 1561 attattcaagtttacaacaataaagagaaatgttgcctataaaaaaacaatttttggtt 1620  
Db 1561 attattcaagtttacaacaataaagagaaatgttgcctataaaaaaacaatttttggtt 1620  
Qy 1621 tctaagaaggaacggttacgattatgacaaactttggcgccgcaatttc 1669  
Db 1621 tctaagaaggaacggttacgattatgacaaactttggcgccgcaatttc 1669

Db 1621 tctaagaaggacgttacgattatgagcaactttggcgccgcgaattc 1669

RESULT 2

AAx78002

ID AAX78002 standard; cDNA; 1737 BP.

XX AC AAX78002;

XX AC AC

DT 19-AUG-1999 (first entry)

XX DE Tobacco GntI cDNA.

XX GntI; beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;

KW deficient; defective; detection; transgenic plant; sugar residue;

KW medicine; antigenicity; deglycosylation; tobacco; ds.

XX OS Nicotiana tabacum.

XX FH Key

FT CDS 127..1467

FT /\*tag= a

FT /product= "GntI"

PN DE19754622-A1.

XX 10-JUN-1999.

XX 09-DEC-1997; 97DE-1054622.

XX 09-DEC-1997; 97DE-1054622.

XX (VSCH/) VON SCHAEWEN A.

XX Von Schaeuwen A;

XX WPI: 1999-338905/29.

DR P-PSDB; AAY08889.

XX Nucleic acid encoding plant N-acetylglucosaminyl transferase I

PT useful for generating plants producing glycosylated proteins of low

PT antigenicity

XX Claim 7; Page 22-24; 37pp; German.

XX This invention describes novel plant N-acetylglucosaminyl transferase I

CC (GntI) proteins and their encoding nucleic acids. The nucleic acid

CC sequences of the invention may be used for recombinant production of the

CC encoded proteins, which are then used to raise antibodies (Ab) for

CC identifying plants with deficient or defective production of the protein.

CC They may also be used to detect such plants by hybridization and to

CC isolate related sequences from other plants or to generate antisense or

CC sense constructs for reducing/deleting GntI protein activity in plants.

CC These transgenic plants may be used to produce glycoproteins with

CC minimal, uniform and defined sugar residues. Such glycoproteins are

CC useful in medicine and research, e.g. human glucocerebrosidase for

CC treating Gaucher's disease. Plants which are defective or deficient in

CC production of the GntI protein can be made to produce glycoproteins with

CC minimal, uniform and defined sugar residues, of low antigenicity. Use of

CC these plants eliminates the need for the difficult isolation and

CC deglycosylation of native proteins or preparation in defective animal

CC cells. This sequence represents tobacco (Nicotiana tabacum) GntI encoding

CC cDNA.

XX Sequence 1737 BP; 510 A; 337 C; 390 G; 500 T; 0 other;

Query Match

Best Local Similarity 75.2%; Score 1255; DB 20; Length 1737;

Matches 1412; Conservative 0; Mismatches 140; Indels 33; Gaps 4;

Qy 52 gataagagggaacagtttgcgttattgattacgggtacattctcgtggtcgtctcgc 111

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Db 126 gatgagagggaacaaagttttgctgctgatttcgcgtacactctcctctcttctggtcgtcgcg 185

Qy 112 cttcatctacacagatgcgggtttttcgcgcacacagtcagaaatattgtagaccgccttgc 171

|||||

Db 186 cttcatctacacagatgcgggtttttcgcgcacacagtcagaaatattgcagatcgccttgc 245

|||||

Qy 172 tgcgtcaattgaagcagaaaaatcattgtacaagtcacaccagattgcttattgacaagat 231

|||||

Db 246 tgcgtcaattgaagcagaaaaatcattgtacaagtcacaccagattgcttattgaccagat 305

|||||

Qy 232 tagccagcagcaagggaagatgtagctcttgaagaacaaatgaagcatcagaccacgga 291

|||||

Db 306 tagccgcagcagcaagggaagatgtagctcttgaagaacaaatgaagcgtcagaccacgga 365

|||||

Qy 292 gtcccggaattaaaggccttcttcaggatctctgaagtaagggcataaaaaagttta 351

|||||

Db 366 gtccgcacaattaaaggccttcttcaggatctctgaagtaagggcataaaaaagttgat 425

|||||

Qy 352 cggagatgtgcagatccagtcggcagctgtagttgttattgcttgcagtcgactaacta 411

|||||

Db 426 cggaaatgtacagatccagtcggcgtgctgtagttgttattgcttgcagtcggctgatta 485

|||||

Qy 412 cctggagaggaactattaaatccatctttaaataccaaacatctgttgcatcaaaatacc 471

|||||

Db 486 cctggaaagaaactattaaatccatctttaaataccaaatctgttgcgcaaaatacc 545

|||||

Qy 472 tcttttcataccagagatggatcaaatcctgatgtgaagaaagcttgccttgagctatgg 531

|||||

Db 546 tcttttcataccagagatggatcaaatcctgatgtcagaaagcttgccttgagctatga 605

|||||

Qy 532 tcagctgacgtatattgcagcaacttggattatgaacctgtgcatactgaaagaccagggga 591

|||||

Db 606 tcagctgacgtatattgcagcaacttggattatgaacctgtgcatactgaaagaccagggga 665

|||||

Qy 592 actggttcataactacaagattgcacgtcattacaaagtggcattggatcagctgtttca 651

|||||

Db 666 gctgattgcatactacaaaattgcacgtcattacaaagtggcattggatcagctgtttca 725

|||||

Qy 652 caagcataattttgacgtgttatacatacagaagatgatagaaattgcctgctgattt 711

|||||

Db 726 caagcataattttgacgtgttatacatacagaagatgatagaaattgcctgctgattt 785

|||||

Qy 712 ttttgactattttgaggtggagctactctcttgacagagacaaagtcgattatggctat 771

|||||

Db 786 ttttgactattttgaggtggagctactctcttgacagagacaaagtcgattatggctat 845

|||||

Qy 772 tctctcttggaatgacaatggcaaaaggcagttcgtccaaagatcctgatgctctttaccg 831

|||||

Db 846 tctctcttggaatgacaatggcaaaaggcagtttgcctcaagatccttatgctctttaccg 905

|||||

Qy 832 ctacagactttttctcctggtcttggatggatgcttccaaaatcaacttgcgcgaactatc 891

|||||

Db 906 ctacagactttttctcctggtcttggatggatgcttccaaaatcaacttgcgcgaactatc 965

|||||

Qy 892 tccaaagtggcacaaggcttactgggatgactggtgactggtgaaagaaatccacagagg 951

|||||

Db 966 tccaaagtggcacaaggcttactgggacagctggtgaagactcaagagaaatccacagagg 1025

|||||

Qy 952 tcgacaaattttcgcgcagagatttgacagacgtacacatttttggtgagcatggtttcag 1011

|||||

Db 1026 tcgacaaattttcgcgcagagatttgacagacataataattttggtgagcatggtttcag 1085

|||||

Qy 1012 tttggggcagttttttaagcagatattcttgagccaatttaagctaaatgatgtccaggttga 1071

|||||

Db 1086 tttggggcagttttttaagcagatattcttgagccaatttaagctaaatgatgtccaggttga 1145

|||||

Qy 1072 ttggaagtcaatggacctaaagtattacccttttgaggagacaactatgtgaacactttggcga 1131

|||||

Db 1146 ttggaagtcaatggaccttagttacccttttgaggagacaatttacgtgaaacactttggtga 1205

|||||

Qy 1132 cttggttataaaaggctaaagccatccacggagctgactgtcttttgaagcatttaacat 1191

|||||

Db 1206 cttggttataaaaggctaaagccatccatggagctgactgtcttcttgaagcatttaacat 1265

|||||

QY	1192	agatggtgatgtcgttattcagttacagagaccacactagactttgaagatactcgtcgaca	1251	PR	04-MAY-1999;	99US-0132484.
				PR	05-MAY-1999;	99US-0132485.
Db	1266	agatggtgatgtcgttattcagttacagagatcaactagactttgaaatactcgacggca	1325	PR	06-MAY-1999;	99US-0132486.
				PR	07-MAY-1999;	99US-0132487.
QY	1252	gtttgcatTTTTgaagaatggaagatggtgtaccacgggcagcatataaauggatagt	1311	PR	11-MAY-1999;	99US-0132863.
				PR	14-MAY-1999;	99US-0134256.
Db	1326	atttggcatTTTTgaagaaTgggaagatgggtgtaccacgtgcagcatataaaaggaatagt	1385	PR	14-MAY-1999;	99US-0134218.
				PR	14-MAY-1999;	99US-0134219.
QY	1312	agttttccggtttcaaacatctagacgtgtgttccctgtttccctgattctcttcgcaca	1371	PR	14-MAY-1999;	99US-0134221.
				PR	14-MAY-1999;	99US-0134370.
Db	1386	agttttccggttaccacacgTccagacgtgtattccctgtttggccatgattcgtttcaaca	1445	PR	18-MAY-1999;	99US-0134768.
				PR	19-MAY-1999;	99US-0134941.
QY	1372	acttgaagtTgaagatactttagcgaagatatgatt---ggagcctgagacaacaatttga	1428	PR	20-MAY-1999;	99US-0135124.
				PR	21-MAY-1999;	99US-0135353.
Db	1446	actcgaattTgaagatactattcaacagagatatgattcgaggccggcgaatttttga	1505	PR	24-MAY-1999;	99US-0135629.
				PR	25-MAY-1999;	99US-0136021.
QY	1429	cttatttggtagatacatcttTgaagagctgcacacgaaagtatgactaccagtagactac	1488	PR	27-MAY-1999;	99US-0136392.
				PR	28-MAY-1999;	99US-0136782.
Db	1506	cttatttgggtaggatgcctgcagctgcacactaaacccatgattt-----accagttac	1558	PR	01-JUN-1999;	99US-0137222.
				PR	03-JUN-1999;	99US-0137528.
QY	1489	atcacaatttttaattTtaattggaagaaacccactgct-----t	1527	PR	04-JUN-1999;	99US-0137502.
				PR	07-JUN-1999;	99US-0137724.
Db	1559	atacaacgttttaattTtatacggaggagctcactgttctagtTtgaagggatatcggc	1618	PR	08-JUN-1999;	99US-0138094.
				PR	10-JUN-1999;	99US-0138540.
QY	1528	attgttggaaTggtgaatacatcaccacatcctattat--ttcaagtttacaacataaag	1585	PR	10-JUN-1999;	99US-0138847.
				PR	14-JUN-1999;	99US-0139119.
Db	1619	ttcttagtatTggtgaatacatcacaacacattatttttaagtTtccagaacataaag	1678	PR	16-JUN-1999;	99US-0139452.
				PR	17-JUN-1999;	99US-0139453.
QY	1586	aggaaatgtTgcctctataaaacaa	1610	PR	18-JUN-1999;	99US-0139454.
				PR	18-JUN-1999;	99US-0139455.
Db	1679	aggaaatgtgacctgttaaagacta	1703	PR	18-JUN-1999;	99US-0139456.
				PR	18-JUN-1999;	99US-0139457.
				PR	18-JUN-1999;	99US-0139458.
				PR	18-JUN-1999;	99US-0139459.
				PR	18-JUN-1999;	99US-0139460.
				PR	18-JUN-1999;	99US-0139461.
				PR	18-JUN-1999;	99US-0139462.
				PR	18-JUN-1999;	99US-0139463.
				PR	18-JUN-1999;	99US-0139463.
				PR	18-JUN-1999;	99US-0139750.
				PR	18-JUN-1999;	99US-0139763.
				PR	21-JUN-1999;	99US-0139817.
				PR	22-JUN-1999;	99US-0139899.
				PR	23-JUN-1999;	99US-0140353.
				PR	23-JUN-1999;	99US-0140354.
				PR	24-JUN-1999;	99US-0140695.
				PR	28-JUN-1999;	99US-0140823.
				PR	29-JUN-1999;	99US-0140991.
				PR	30-JUN-1999;	99US-0141287.
				PR	01-JUL-1999;	99US-0141842.
				PR	01-JUL-1999;	99US-0142154.
				PR	02-JUL-1999;	99US-0142055.
				PR	06-JUL-1999;	99US-0142390.
				PR	08-JUL-1999;	99US-0142803.
				PR	09-JUL-1999;	99US-0142920.
				PR	12-JUL-1999;	99US-0142977.
				PR	13-JUL-1999;	99US-0143542.
				PR	14-JUL-1999;	99US-0143624.
				PR	15-JUL-1999;	99US-0144005.
				PR	16-JUL-1999;	99US-0144085.
				PR	16-JUL-1999;	99US-0144086.
				PR	19-JUL-1999;	99US-0144325.
				PR	19-JUL-1999;	99US-0144331.
				PR	19-JUL-1999;	99US-0144332.
				PR	19-JUL-1999;	99US-0144333.
				PR	19-JUL-1999;	99US-0144334.
				PR	19-JUL-1999;	99US-0144335.
				PR	20-JUL-1999;	99US-0144352.
				PR	20-JUL-1999;	99US-0144632.
				PR	21-JUL-1999;	99US-0144884.
				PR	21-JUL-1999;	99US-0144814.
				PR	21-JUL-1999;	99US-0145086.
				PR	21-JUL-1999;	99US-0145088.

RESULT 3

AAC39540

ID AAC39540 standard; DNA; 1639 BP.

XX

AC AAC39540;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 25006.

XX

KW Hybridisation assay; genetic mapping; gene expression control;

KW

KW protein identification; signal transduction pathway;

KW

metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

PF

PF 25-FEB-2000; 2000EP-0301439.

XX

XX 25-FEB-1999; 99US-0121825.

PR

PR 05-MAR-1999; 99US-0123180.

PR

PR 09-MAR-1999; 99US-0123548.

PR

PR 23-MAR-1999; 99US-0125788.

PR

PR 25-MAR-1999; 99US-0126264.

PR

PR 29-MAR-1999; 99US-0126785.

PR

PR 01-APR-1999; 99US-0127462.

PR

PR 06-APR-1999; 99US-0128234.

PR

PR 16-APR-1999; 99US-0128714.

PR

PR 19-APR-1999; 99US-0129845.

PR

PR 21-APR-1999; 99US-0130077.

PR

PR 23-APR-1999; 99US-0130449.

PR

PR 23-APR-1999; 99US-0130510.

PR

PR 28-APR-1999; 99US-0130891.

PR

PR 30-APR-1999; 99US-0131449.

PR

PR 30-APR-1999; 99US-0132048.

PR

PR 30-APR-1999; 99US-0132407.



Qy	845	ctcggcttggatggatggttttcaaaatcaacttggctcgaactatctccaagtgcca	904	PR	16-APR-1999;	99US-0129845.
Db	818	ctcggcttggatggatgctcaagatcgacttggatgagttatccaccaagtgcca	877	PR	19-APR-1999;	99US-0130077.
Qy	905	aaggcttactggatgactggctaaagctgaaagaaaatcacagaggtgcacaattatt	964	PR	21-APR-1999;	99US-0130449.
Db	878	aaggcttactggatgattggctgagactaaagaaaaccataaagccgcgaattcatt	937	PR	23-APR-1999;	99US-0130510.
Qy	965	cgcccaagaagtgttcagaaacgtacaattttggtgagcatggtctagtttggggcagttt	1024	PR	28-APR-1999;	99US-0130891.
Db	938	cgaccggaagtctgtagaacatacaaatittggtgaactgggtctagtttgggacagttt	997	PR	30-APR-1999;	99US-0131449.
Qy	1025	tttaagcagatcttgagccaattaaagctaaatgatgtccagggttgattggaagcaatg	1084	PR	99US-0132048.	99US-0132407.
Db	998	ttcagtcagatcttgaaacctataaagctaaacgatgtgacggttgactggaaagcaaa	1057	PR	99US-0132407.	99US-0132484.
Qy	1085	gacctaaagttaacttttggagcaacctatgtgaacactttgacacttggtaaaaaag	1144	PR	04-MAY-1999;	99US-0132485.
Db	1058	gacctgggataacctgacagagggaacctataccaagtactttctggttagtgagaca	1117	PR	05-MAY-1999;	99US-0132486.
Qy	1145	gctaagcccatccacggagctgctgtgttttgaagcatttaacatagatggtgatgtg	1204	PR	06-MAY-1999;	99US-0132487.
Db	1118	gcacgaccaattcgaaggttctgaccttgccttaaggctcaaaacataaaagatgatgtt	1177	PR	07-MAY-1999;	99US-0132487.
Qy	1205	cgtattcagtcagagacccaactagactttgaagatatacgtcgcacagtttggcatttt	1264	PR	11-MAY-1999;	99US-0132863.
Db	1178	cgtaccggtataaagaccaagttagatgttgaacgttttgaacgtttgacggtttggtatatt	1237	PR	99US-0134256.	99US-0134219.
Qy	1265	gaagaatggaaggatggtgtaccacggcgagcatataaaggatagtagtttccggttt	1324	PR	99US-0134219.	99US-0134221.
Db	1238	gaagaatggaaggatggtgtccctcgacacagcatataaaggagtagtgggttcgaatc	1297	PR	14-MAY-1999;	99US-0134370.
Qy	1325	caaacatcagacgtgttccctgtttccctgattctcttcgacaaacttggagttgaa	1384	PR	18-MAY-1999;	99US-0134768.
Db	1298	cagacaacagacgtgttccctgtttccctgattctcttcgacaaacttggagttgaa	1357	PR	19-MAY-1999;	99US-0134941.
Qy	1385	gatacttagcgaaga	1399	PR	20-MAY-1999;	99US-0135124.
Db	1358	aattcctgaigcaaa	1372	PR	99US-0135353.	99US-0136021.
RESULT	4			PR	27-MAY-1999;	99US-0136392.
AAC50197				PR	28-MAY-1999;	99US-0136782.
ID	AAC50197	standard; DNA; 1641 BP.		PR	03-JUN-1999;	99US-0137222.
XX	AC	AAC50197;		PR	99US-0137528.	99US-0137502.
XX	DT	18-OCT-2000 (first entry)		PR	07-JUN-1999;	99US-0137724.
XX	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 63936.		PR	08-JUN-1999;	99US-0138094.
XX	KW	Hybridisation assay; genetic mapping; gene expression control;		PR	10-JUN-1999;	99US-0138540.
XX	KW	protein identification; signal transduction pathway;		PR	14-JUN-1999;	99US-0138847.
XX	XX	metabolic pathway; promoter; termination sequence; ss.		PR	16-JUN-1999;	99US-0139119.
OS	Arabidopsis thaliana.			PR	17-JUN-1999;	99US-0139452.
XX	PN	EP1033405-A2.		PR	16-JUN-1999;	99US-0139453.
XX	PD	06-SEP-2000.		PR	17-JUN-1999;	99US-0139452.
XX	PF	25-FEB-2000; 2000EP-0301439.		PR	18-JUN-1999;	99US-0139455.
XX	XX	25-FEB-1999; 99US-0121825.		PR	18-JUN-1999;	99US-0139456.
PR	05-MAR-1999;	99US-0123180.		PR	99US-0139457.	99US-0139458.
PR	09-MAR-1999;	99US-0123548.		PR	99US-0139458.	99US-0139459.
PR	23-MAR-1999;	99US-0125788.		PR	99US-0139460.	99US-0139461.
PR	25-MAR-1999;	99US-0126264.		PR	99US-0139462.	99US-0139463.
PR	29-MAR-1999;	99US-0126785.		PR	99US-0139750.	99US-0139763.
PR	01-APR-1999;	99US-0127462.		PR	99US-0139817.	99US-0139899.
PR	06-APR-1999;	99US-0128234.		PR	99US-0140353.	99US-0140354.
PR	08-APR-1999;	99US-0128714.		PR	99US-0140695.	99US-0140823.
PR	25-FEB-1999;	99US-0121825.		PR	99US-0140991.	99US-0141287.
PR	05-MAR-1999;	99US-0123180.		PR	99US-0141287.	99US-0142154.
PR	09-MAR-1999;	99US-0123548.		PR	99US-0142055.	99US-0142390.
PR	23-MAR-1999;	99US-0125788.		PR	99US-0142803.	99US-0142920.
PR	25-MAR-1999;	99US-0126264.		PR	99US-0142977.	99US-0143542.
PR	29-MAR-1999;	99US-0126785.		PR	99US-0143624.	99US-0144005.
PR	01-APR-1999;	99US-0127462.		PR	99US-0144085.	99US-0144086.
PR	06-APR-1999;	99US-0128234.		PR	99US-0144325.	99US-0144331.
PR	08-APR-1999;	99US-0128714.		PR	99US-0144332.	99US-0144333.



Qy	722	tttgaggctggagctactcttcttgacagacaaagtcgattatggctatttctcttgg	781
Db	696	tttgaggctggagctagctctcaaggatagggaataaaaccattatgctgttcacatg	755
Qy	782	aatgacaatggacaaaggcagttcgtccaaagatcctgctcttttaacgcctgcagcttt	841
Db	756	aatgataatggacaaagcagtttgcgtgatccctatgcgtataaccgacagatttt	815
Qy	842	tttctggcttggatggatgctttcaaaatcaacttgggtccgaactatctccaaagtgg	901
Db	816	tttctggccttgggtggatgctccaagagatcgacttggatgagtatccccaagtgg	875
Qy	902	ccaaagg-cttactgggagctcgtggctgaaggctgaaagaaaatcacagaggtcgacaatt	960
Db	876	ccaaaggccttactgggagctgattggctggagactaaaggaaaaccataaaggccgaatt	935
Qy	961	tattcgcccaagatttgcaagaacgtacaatttttggctgagcaatggctctagtttgggca	1020
Db	936	cattcgaccgggaagctgtgagaacacacaaattttggtgaacatgggtctagtttggaca	995
Qy	1021	gtttttlaagcagttactgtgagccaattaagctaaatgatgtccaggtgtgtaggaagtc	1080
Db	996	gtttttcagtcagtatctggaaacctataaagctaaacgatgtgacggttgactggaaagc	1055
Qy	1081	aatggacctaaagttaacctttggagagacaactatgtgaacacctttggcgacttggtaa	1140
Db	1056	aaaggacctgggaatacctgacagagaggaacataccaagtaactttcttggcttagtga	1115
Qy	1141	aaagctaaagccatccagcgagctgactgctgtttgaaagcatttaacatagatggtga	1200
Db	1116	acagcgacccaattccaaggttctgacctgtctgaagctcttaaggtccaaacataaaaggatga	1175
Qy	1201	tgtcgctattcagtcacagagaccaactagacttttgaagatatacgtctcgacagttggcat	1260
Db	1176	tgttcgtatccgtataaagaccacagtagattgtgaacgcatctgcaggggaatttggat	1235
Qy	1261	tttgaagaatggaagatggtgtaccgcggcgacatataaaagggatagtagtittccg	1320
Db	1236	atttgaagaatggaagatggtgtgctcgacagcatataaaaggagtagtgggtttcg	1295
Qy	1321	gtttcaaacatctagacgtgtgtctctgtttccctgatctcttcgacaacttggagt	1380
Db	1296	aatccagacaaagacgtgtattctctgttggccagattcttgaatgcagcttggat	1355
Qy	1381	tgaagatacttagcgaaga	1399
Db	1356	tcgaaattctgatgcata	1374
RESULT 5			
AAx78003			
ID	AAx78003 standard; cDNA; 510 BP.		
AC	AAx78003;		
XX			
DT	19-AUG-1999 (first entry)		
XX			
DE	A. thaliana GnTI-homologue cDNA fragment.		
XX			
KW	GnTI; beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;		
KW	deficient; defective; detection; transgenic plant; sugar residue;		
KW	medicine; antigenicity; deglycosylation; ds.		
OS	Arabidopsis thaliana.		
XX			
PN	DE19754622-A1.		
XX			
PD	10-JUN-1999.		
XX			
PF	09-DEC-1997; 97DE-1054622.		
XX			
PR	09-DEC-1997; 97DE-1054622.		

Db	11		11		1		481	cgatcagattcttccctggccacggctgg	510
Db	947	gactttttccctggccttggatggcgtgttggctgattctctgggcagaaactagagccc	1000						
QY	896	aagtggccaaaggcttacttggtggtgactgactggttaaggctgaaagaaaaatcacagaggtcga	955						
Db	1007	aagtggccaaaggcttctgggacgactggtatgctgagacacctgagcagcggaaggggcgg	1066						
QY	956	caattttattcgcacgaagtttgcagaacgtacaaatttttgggtgagcatggttcttagtttg	1015						
Db	1067	gcttgatttcgccaagaatttcaagaactatgacatttggctgcaagggttgagccat	1126						
QY	1016	ggcgagtttttaagcagtatcttgagccaattaaagctaaatgatgtccaggttgattgg	1075						
Db	1127	ggcgagttctttgaccagcatcttaaatctcaagctgaaccgacgttcgtgcccttc	1186						
QY	1076	aagtcaatggaccctaagtattaccttttgaggacaactatgtgaaacacttttgacacttg	1135						
Db	1187	accagttggacctgtctacctgcaggcgggaggtcctatgacccgggacttctctccacg	1246						
QY	1136	gttaaaaggctaaagcccatccacggagctgctgctgttttgaagcatttaacatagat	1195						
Db	1247	gtctatggtgcccccagctgcaggtggagaaagtggagccaatgacggaagaaactg	1306						
QY	1196	ggtgatgtgcgtatttcagtagacagacaaactagactttgaagatatcgctgcacagttt	1255						
Db	1307	ggggaggtgcgggtacagcacactagcagcagctttaaaggccttgcgaagccctg	1366						
QY	1256	ggcatttttgaagaatggaaggtggtaccacgggcagcatataaaagggtatagtt	1315						
Db	1367	ggtgtcatggcgcacctcaagtcgtgttcccgagctggtctaccctggcattgtcact	1426						
QY	1316	tcccggtttc	1325						
Db	1427	tccagttccc	1436						
RESULT	7								
AAQ25594									
ID	AAQ25594	standard; CDNA; 2485 BP.							
XX	AAQ25594;								
AC	AAQ25594;								
XX	03-DEC-1992	(first entry)							
DT									
DE									
XX									
XX									
KW									
KW									
KW									
OS									
OS									
XX									
FH									
FT									
FT									
FT									
FT									
FT									
XX									
PN									
XX									
PD									
XX									
PF									
XX									
PR									
XX									
PA									
PI									
XX									
DR									
XX									

Db	481	cgatcagattcttccctggccacggctgg	510						
Db	947	gactttttccctggccttggatggcgtgttggctgattctctgggcagaaactagagccc	1000						
QY	896	aagtggccaaaggcttacttggtggtgactgactggttaaggctgaaagaaaaatcacagaggtcga	955						
Db	1007	aagtggccaaaggcttctgggacgactggtatgctgagacacctgagcagcggaaggggcgg	1066						
QY	956	caattttattcgcacgaagtttgcagaacgtacaaatttttgggtgagcatggttcttagtttg	1015						
Db	1067	gcttgatttcgccaagaatttcaagaactatgacatttggctgcaagggttgagccat	1126						
QY	1016	ggcgagtttttaagcagtatcttgagccaattaaagctaaatgatgtccaggttgattgg	1075						
Db	1127	ggcgagttctttgaccagcatcttaaatctcaagctgaaccgacgttcgtgcccttc	1186						
QY	1076	aagtcaatggaccctaagtattaccttttgaggacaactatgtgaaacacttttgacacttg	1135						
Db	1187	accagttggacctgtctacctgcaggcgggaggtcctatgacccgggacttctctccacg	1246						
QY	1136	gttaaaaggctaaag							



Human and rabbit DNA sequences encoding GNT I enzymes - for converting mannose to hybrid and complex N-glycan(s)

Claim 3; Fig 4; 57pp; English.

Rabbit liver Gnt I was purified and digested with pepsin, then trypsin. The peptide fragments were sequenced and used to design degenerate PCR primers (see AAQ30220-5). cDNA was prepared from total RNA from rabbit liver. PCR was carried out on the cDNA preparation. One of the two PCR products was cloned into the SmaI site of pGEM7z for sequencing and then used as a riboprobe. The riboprobe was used to screen a rabbit liver cDNA library in lambda gt10. The largest insert in a positive clone was 1.6kb. An 80bp riboprobe was prepared from the 5'-terminal of the 1.6kb insert and used to rescreen the library. The largest cDNA insert was cloned into pGEM-7z to obtain pGEM-7z-rcgntI. The full-length rabbit Gnt I coding sequence was eventually obtained from overlapping clones. See also AAQ25595.

Sequence 2485 BP; 549 A; 623 C; 644 G; 669 T; 0 other:

Query Match 10.9%; Score 181.6; DB 13; Length 2485;  
Best Local Similarity 54.18; Pred. No. 2.4e-43;

11

QY	602	tactacaagatgacgctcattacaagtgggcattggtacagctgtttccacaagcataat	661
Db	599	tactacaagatgcgcgcgcttacgcctgggcattcggccaaatcttccaaatttccaac	658
QY	662	tttagccgtgtatcatcatagaagatgatatggaaaattgcgtgatttttttgaactat	721
Db	659	taccacagcagctgtggtggtggaagatgattctcgagggtggccacgagactcttttgaglac	718
QY	722	tttgaggctggagctactcttttcagacagacgaagtcgattatggctattctctcttgg	781
Db	719	ttcagggccacttaccacactgttgaaagcagaccctccctcgtgtgtgctgcctcgg	778
QY	782	aatgacaaatggacaaggcaggtctgtcca-----agatccctgagctctttaccgcgtca	835
Db	779	aatgacaaatggcaagaacagatggtagactcgtagtaagccaagattactctaccgcgaca	838
QY	836	gactttttctcgttctggatggatgtcttcacaaatcaacttggctcgcaactatctcca	895
Db	839	gattctttctcgttctggcttggctgttactgttggtgaaactctggcgaagggctgagccc	898
QY	896	aagtgcccaagggcttactgggatgactggctaaagctgaaagaaaaatccacagaggtcca	955
Db	899	aagtgcccaagcctctcgggatgactggatgcgccgcctgagcagcgaagggggagg	958
QY	956	caatttattcgcgcagaagtttgcagaagctacaaattttggtgagcatggttctagtttg	1015
Db	959	gcctgtgtgcgtccagaatctcaagaacaatgacatttggcggaaagggctgagccat	1018
QY	1016	gggcagttttttaaagcagtatctttgagccaaatgaagctaaatgatgtccaggttgattgg	1075
Db	1019	gggcagttctttgaccagcatctcaagttcatcaagctggaaccgagcttgtacccttc	1078
QY	1076	aagtcacatggacctaaagttaaccttttgaggacaactatgtgaacacctttggcgacttg	1135
Db	1079	accgcagctggacctgtcactcttcagcaggagactatgacgggatttctctgtcgt	1138
QY	1136	gttaaaaaggctaaagcccatccacggagctgatcgtgttttgaaagcaatttaacatagat	1195
Db	1139	gtttatgtgtctcccagttacaggttgagaagatggaggaccaatgaccggaaaggacta	1198
QY	1196	gggtgatgtgcgtatttcagtcagagaccaactagactttgaagatatcgtctcgacagttt	1255
Db	1199	ggagagggtgcgctacagtacacagggcagggaacagcttcaaggctttcgccaagccctg	1258
QY	1256	ggcatttttgagaatggaaggatgggttaccacggcgacacataaaggatagtagtt	1315
Db	1259	ggtgtcatggtgatccttcaaatcaggtgtgcccaagggctggaaccgggcatgtgtcacc	1318

Qy 1316 ttccggttc 1325  
||| |||  
Db 1319 ttctattcc 1328

## RESULT 8

AAI69837  
ID AAI69837 standard; cDNA; 1335 BP.

AC AA169837;

DT 19-DEC-2001 (first entry)

DE Human beta-1,2-N-acetylglucosaminyltransferase I cDNA.

Human; beta-1,2-N-acetylglucosaminyltransferase; GnT; GnTI;  
KW  
maltose-binding protein; MBP; fusion; sugar chain synthesis  
KW

OS Homo sapiens.

PN JP2001178453-A.

PD 03-JUL-2001.

24-DEC-1999; 99JP-0368127.

PR 24-DEC-1999; 99JP-0368127.

РА (ТОУМ ) ТОУОВО КК.

FA XX  
(FUJI/ ) FUJIYAMA K.

DR WPI; 2001-599903/68.  
DR P-PSDB: AAG67094.

AA  
PT  
PT  
PT

preparation of maltose-binding protein-N-acetylglucosaminyltransferase for use in sugar chain synthesis, comprises using *Escherichia coli* -

PS Claim 5; Page 9-11; 14pp; Japanese.

The invention relates to the preparation of a maltose-binding protein (MBP)-N-acetylglucosaminyltransferase (Gnt) fused protein. The method comprises transforming *Escherichia coli* with an expression vector encoding the fusion protein. Gnt antibody is useful in sugar chain synthesis and in a plasmid pMAL-c2/EcoRI-XbaI fragment to give a new plasmid pMGNT-1. The present sequence encodes human Gnt.

SQ Sequence 1335 BP; 215 A; 428 C; 441 G; 251 T; 0 other;

Query Watch	9.9%	Score 165.6	DB 22	Length 1335
Best Local Similarity	52.7%	Fred. No. 1e-38		
Matches 385:	Conservative	0	Mismatches 339	Indels 6
			Gaps 6	Gaps 1

Qy 602 tactacaagattgcacgctcattacaagtcacgctgatttcacaagcataat 661

Db 544 tactacaagatcgcgcgcgcactaccgcctgggcgcctgggcagggtcttccggcagtttcgc 603

Qy 662 tttaqccqtgttatcatactagaagatgatatggaatttctgctgaatttttttgaactat 771

Db 604 ttccccqccqccgtggtggtggaggatgacctggaggtagccccggaactcttccgaat ac 663

Qy 722 tttagqctggagctactctcttgacagagacaaatcgattatcactatctctctcg 781

Db

664 ttctggggccacctatccgctgctgaagccgacccctccctatgcatccatctcggcctcg 733

[illegible][illegible][illegible][illegible]

QY 896 aagtgccaaaggcttactggatggctgaagcgtgaaagaaaaacacagaggtcga 955  
|||||  
Db 844 aagtgccaaaggcttctggacgactggatgcgcgccgagcagcgagggcg 903  
QY 956 caattatttcgcgaagtttgcagaacgtacaaattttggtgagcatggttctagt 1015  
|||||  
Db 904 gctgcatacgccttgagatctcagaacgtatgacctttggccgaagggtgtgagccac 963  
QY 1016 ggcagtttttaagcagtattcttgagccaattagctaaatgatgtccaggttgattgg 1075  
|||||  
Db 964 ggcagttcttgaccacgaccccaagtttatcaagctgaaccagcagtttgcacttc 1023  
QY 1076 aagtcataagcctaagtttaacctttttgagagacaactatgtgaacacttttggcacttg 1135  
|||||  
Db 1024 acccagctggactgtcttacctgcagcgaggagcctatgaccagatttctcccccgc 1083  
QY 1136 gttaaaaaggctaaagccatccacgagctgatgtgtgttttgaagcatttaacatagat 1195  
|||||  
Db 1084 gtcacaggtgctcccccagctgcaggtggagaaagtggagaccatgaccggaaggactg 1143  
QY 1196 ggtgatgtgctattcagtcacagagacccaactagactttgaaagatatgcctgcagattt 1255  
|||||  
Db 1144 gggaggtgcgggtgcagtatacggcagggagacgttcaaggctttccccaaggctctg 1203  
QY 1256 ggcatttttgaagaatggaagatggtgtaccacggcagcatataaaaggatagtt 1315  
|||||  
Db 1204 ggtgatggatgaccttaagtcagggggtctccgagagctggctaccggggtattgtcacc 1263  
QY 1316 ttccgggttc 1325  
|||||  
Db 1264 ttccagttcc 1273

RESULT 9

AAQ25595  
ID AAQ25595 standard; cDNA; 3230 BP.  
XX  
AC AAQ25595;  
XX  
DT 03-DEC-1992 (first entry)  
DE  
DE Human Gnt I cDNA clone.  
XX  
XX High mannose glycoprotein; UDP-N-acetyl-glucosamine;  
KW alpha-3-D-mannoside; beta-1,2-N-acetyl-glucosaminyl transferase 1;  
KW N-glycans; chronic myeloid leukaemia; ss.  
XX  
OS Homo sapiens.

XX  
XX  
FH Key Location/Qualifiers  
FT CDS 903..2240  
FT /\*tag= a  
FT /product= human\_Gnt\_I  
XX

PN W09209694-A.  
XX  
PD 11-JUN-1992.  
XX  
PF 29-NOV-1991; 91WO-CA00417.  
XX  
PR 30-NOV-1990; 90US-0620098.  
XX

XX (HSCR-) HSC RES & DEV LP.  
XX  
XX Sarkar M, Schachter H;  
XX WPI; 1992-217073/26.  
XX P-PSDB; AAR24781.  
XX  
XX Human and rabbit DNA sequences encoding Gnt I enzymes - for  
XX converting mannose to hybrid and complex N-glycan(s)  
XX

Claim 12; Fig 6; 57pp; English.

PS  
XX A rabbit Gnt I cDNA probe was prepared using PCR on rabbit liver  
CC cDNA. The probe was used to screen an amplified human genomic DNA  
CC library in lambda EMBL3 prepared from chromosomal DNA of chronic  
CC myeloid leukemia cells. The coding sequence was located in a 4.0kb  
CC fragment of human genomic DNA by screening with a probe containing  
CC the entire coding region of the rabbit Gnt I cDNA. The fragment was  
CC sequenced and found to contain an ORF coding for a protein of 445  
CC amino acids (2 less than the rabbit enzyme). The similarity between  
CC the human and rabbit enzyme coding sequences is 85%.  
XX See also Q25594.

SQ Sequence 3230 BP; 575 A; 938 C; 913 G; 804 T; 0 other;

Query Match 9.8%; Score 164; DB 13; Length 3230;

Best Local Similarity 52.6%; Pred. No. 5e-38;  
Matches 384; Conservative 0; Mismatches 340; Indels 6; Gaps 1;

QY 602 tactacaagattgcacgtcattacaagtggcgatggatcagctgtttcacagcataat 661  
|||||

Db 1446 tactacaagatgcgcgcactaccgctggcgctggccaggtctccgcagttccg 1505  
|||||

QY 662 tttagccgtgttatcactagaaatgatataaggaaattgctgctgttttttactat 721  
|||

Db 1506 ttccccgcggcgtggtggtgagatgacctggaggtggccccggaacttctcagtac 1565  
|||

QY 722 tttagggctggagctactctttgacagagacaagtcgattatgctatttctcttg 781  
|||||

Db 1566 ttctgggcccacatatccgctgtgaaggccgacctctctgtgctcctggcctgg 1625  
|||||

QY 782 aatgacaatggcaaaagcagcttcgtcccaag-----atcctgatgctctttaccgctca 835  
|||||

Db 1626 aatgacaacggcaaggagcagatggtggacgccagcagcagctgagctctaccgcac 1685  
|||||

QY 836 gactttttctcgtgtgttgatggatgcttttcaaaatacaacttggctcgaactatctca 895  
|||||

Db 1686 gactttttctcgtgtgttgatggatgcttttggccgagctctgtgctgagctggagcc 1745  
|||||

QY 896 aagtgccaaaggcttactggatggatgactgaagcgtgaaagaaaaacacagaggtcga 955  
|||||

Db 1746 aagtgccaaaggcttctggacgactggatgcgcgccggagcagcgagggcg 1805  
|||||

QY 956 caatttttcgccagaagtttgcagaacgtacaatttttggtagcatggttctagt 1015  
|||||

Db 1806 gctgcatacgccttgagatctcaagaacgatgacctttggccgaagggtgtgacgcac 1865  
|||||

QY 1016 gggcagtttttaagcagtatctttgagccaataaagctaaatgatgtccaggttgattgg 1075  
|||||

Db 1866 gggcagttctttgaccagcacctcaagtttatcaagctgaaccagcagtttgtgcacttc 1925  
|||||

QY 1076 aagtcataagcctaagtttaccttttggaggcaacatactgtgaaacacttttggcgacttg 1135  
|||||

Db 1926 acccagctggacctgtcttacctgcagcgaggcctatgaccgagatttctcgcgcgc 1985  
|||||

QY 1136 gttaaaaaggctaagcccatccacgagctgatgtgttttgaagcatttaacatagat 1195  
|||||

Db 1986 gtctacggtgctccccagctgcaggtggagaaagtggagccaatgacccgagagctg 2045  
|||||

QY 1196 ggtgatgtcgtattcagtcacagagaccaactagactttgaaagatacgtcgacagttt 1255  
|||||

Db 2046 ggggaggtgcgggtgcagtatacggggagggagacagcttcaaggcttccccaaggctctg 2105  
|||||

QY 1256 ggcatttttgaagaatggaagatggtgtaccacggcgagcatataaaagggagtagtt 1315  
|||||

Db 2106 ggtgttatgtagaccttaagtcgggggttccgagagctggctaccggggtattgtcacc 2165  
|||||

QY 1316 ttccgggttc 1325  
|||||

Db 2166 ttccagttcc 2175  
|||||

```
RESULT 10
AAF93806
ID AAF93806 standard; cDNA; 2250 BP.
XX
AC AAF93806;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human cDNA encoding a membrane or secretory protein clone PSEC0120.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; ss.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI: 2001-093989/11.
DR P-PSDB; AAB88379.
XX
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 1; SEQ ID 125; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 2250 BP; 408 A; 561 C; 586 G; 495 T; 0 other;
```

```
Query Match 8.6%; Score 144.2; DB 22; Length 2250;
Best Local Similarity 51.7%; Pred. No. 3.3e-32;
Matches 357; Conservative 0; Mismatches 328; Indels 6; Gaps 1;

Qy 641 cagctgtttccacagcacaatttttagcgtgttatctactataagatgatgaaatt 700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 caggcttccgcagcttccgtcccccgcgcgtggtggaggatgacctggagggtg 468
```

```
QY 701 gctgctgattttttgactattttgaggtggagctactcttcttgacagacaaagtcg 760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 469 gcccggaacttctcagtaacttccagccacactatccgtctgaagccgacccctcc 528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 761 attatgctatttcttcttggaatgacaaatggacaaagcagttcgtccaagat----- 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 529 ctgtgtgctgtcgtccctgggaatgacaaagcgaagcagatggtgacgcacagcag 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 815 cctgctgctcttaccgcctcagacttttctcgtgtcttggtggtggtctttcaaaatca 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 589 cctgagctctctacgcgacccgacttttctcgtggtggtggtggtctgtgtggccgag 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 875 acttggctccgaactatctccaagtggccaaaggttactgggatgactggctaaagctg 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 649 ctctggctgagctggagcccaagtggccaaagccttctgggacgactggatgcgcg 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 935 aaagaaatcacagaggtgcacaaatttctgccagaagtttgccagaagctacaaattt 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 709 ccggagcagcggcagggcggcctgcatacgccttgagatctcaagaacgatgacctt 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 995 ggtgagcatggttctagtttggggcagtttttaagcagtatctttgagccaatgaagcta 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 769 ggcgcaaggggtgtgagccacgggcagttctttgaccgacactcaagtattatcaagctg 828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1055 aatgatgtccaggttgatgggaagtcaatggaccttaagttaaccttttgaggagaaactat 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 829 aaccagcagttgtgcacttcaccagctggacctgtcttaacctgcgcgggagcctat 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1115 gtgaacacatttggcagcttggtaaaaaggcttaagcccatccacggagctgctgtgt 1174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 889 gaccgagatttctcgcgcgcgtctacggtgctcccccagctcaggtggagaagtgaag 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1175 ttgaagaactttaacatagatggtgatgtgctatttcacagacagacccaactagacttt 1234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 949 accaatgaccggaagaggttggggagggttcgggtgcagtatacagggcgagggacagcttc 1008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1235 gaagatactcgtcgacagtttggcatttttgaagaatggaagatggtgtaccacgggca 1294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1009 aaggcttccgcaaggtcgggtgctcatgagaccttaagtccgggggttcgagagct 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1295 gcataataaggagtagtagtttccgggttc 1325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1069 ggtaccggggtattgtcaccttccagttcc 1099
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
ABL22693
ID ABL22693 standard; DNA; 2367 BP.
XX
AC ABL22693;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19552.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
```





the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2078 BP; 467 A; 582 C; 569 G; 460 T; 0 other;

Query Match 2.6%; Score 44; DB 22; Length 2078;  
Best Local Similarity 50.3%; Pred. No. 0.025;  
Matches 172; Conservative 0; Mismatches 155; Indels 15; Gaps 2;

670 tggatgactctctgacagagacagtcgattggttcttcttctggaatgacaa 789  
674 tgcgtggtctggaagagaccctggacattgctggtgatttttcagttctctgagcca 733  
730 tggagctactctctgacagagacagtcgattggttcttcttctggaatgacaa 789  
734 atccatccactgagagagagtcgacagctgctgctgctgctggaatgacaa 793  
790 tggagagcagctgctcaagatctgctgcttcttaccgctcagacttttctctg 849  
794 ggggtatgaacacacgctgagagaccagcactgctgctgctgagacacatgctctg 853  
850 tcttgatggtgcttctcaaaatcaactgt---gtccgaactatctccaaagtggccaa 906  
854 gctggctgggtgctcaggggtcctgttacaaggagaggtgctgagcccaagtggcctac 913  
907 ggcctac-----tggagactgctgctaaagctgaaagaaatcacagaggtcg 954  
914 accggaagaagctctgggattgggacatgtggatgctgctgacacacgagcgggcg 973  
955 acaattattcgcgcagaaagtgttcagaaagtcacattttgg 996  
974 agagtgcatacctcgaactttcccgatcctaccactttgg 1015

RESULT 15

AA158317  
ID AA158317 standard; cDNA; 2353 BP.

AA158317;

22-OCT-2001 (first entry)

Human polynucleotide SEQ ID NO 520.

Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.  
21-JAN-2000; 2000US-0488725.  
25-APR-2000; 2000US-052317.  
09-JUL-2000; 2000US-0598042.  
19-JUL-2000; 2000US-0620312.  
03-AUG-2000; 2000US-0653450.  
14-SEP-2000; 2000US-0662191.  
19-OCT-2000; 2000US-0693036.  
29-NOV-2000; 2000US-0727344.  
(HYSE-) HYSEQ INC.  
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
WPI; 2001-442253/47.  
P-PSDB; AAM39161.  
Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -  
Claim 1; SEQ ID NO 520; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAM38642-AAAM42213) with nototropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.  
Note: The sequence data for this patent did not form part of the printed specification.

Sequence 2353 BP; 529 A; 651 C; 690 G; 483 T; 0 other;

Query Match 2.6%; Score 44; DB 22; Length 2353;  
Best Local Similarity 50.3%; Pred. No. 0.025;  
Matches 172; Conservative 0; Mismatches 155; Indels 15; Gaps 2;

670 tggatgactctgacagagacagtcgattggttcttcttctggaatgacaa 729  
1316 tgcgtggtctggaagagaccctggacattgctggtgatttttcagttctctgagcca 1375  
730 tggagctactcttctgacagagacagtcgattggttcttcttctggaatgacaa 789  
1376 atccatccactgactggagagagtcgacagcctgctactgctcctgctggaatgacaa 1435  
790 tggacaaaggcagttctccaagatctctgcttcttaccgctcagacttttctctg 849  
1436 ggggtatgaacacacgctgagagaccagcactactgctgctgagacacatgctctg 1495  
850 tcttgatgagctcttcaaaatcaactgt---gtccgaactatctccaaagtggccaa 906  
1496 gctggctgggtgctcagaggtcctgttacaaggagaggtgtgagcccaagtggcctac 1555  
907 ggcctac-----tgggactgctggtcgaagctgaaagaaatcacagaggtcg 954  
1556 accggaagaagctctgggattgggacatgtggatgctgagcctggaacacgagcg 1615  
955 acaattattcgcgcagaaagtgttcagaaagtcacattttgg 996  
1616 agagtgcatacctcgaactttcccgatcctaccactttgg 1657

Wed Aug 14 11:48:20 2002

us-09-591-466c-1.rng

Page 16

Search completed: August 13, 2002, 20:33:01  
Job time: 8309 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 17:48:37 ; Search time 3105.21 Seconds  
(without alignments)  
11247.674 Million cell updates/sec

Title: US-09-591-466c-1

Perfect score: 1669  
Sequence: 1 gaattcgccgcctgaga.....actttggccgcggaattc 1669

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

RESULT 1

A95053

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

A95053  
Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH

A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GNTI)

ACTIVITY

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIRE"

ALIGNMENTS

A95053  
Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH

A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GNTI)

ACTIVITY

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIRE"





|||||  
Db 1441 CATACTTTGAAGAGCTGACAGAAAGTATGACTACCAGTAGCTACATGCAACATTTT 1500  
Qy 1501 aatgttaatgaaggaaacccactgcttatgttggaatgagatgaatcaccacacatct 1560  
Db 1501 AATGTTAATGAAGGAACCCACTGCTTATGTGTGGAATGGATGAATCATCACCACATCT 1560  
Qy 1561 attattcaagtttacaacataaagaggaaatgttgcctctataaaacaaaattttttgt 1620  
Db 1561 ATTATTCAAGTTTACAAACATAAAGAGGAATGTTCCTCTATATAAAACAAATTTTGT 1620  
Qy 1621 tctaagaggaacgttacgattatgagcaactttggcgccgcgaattc 1669  
Db 1621 TCTAAGAAGCAAGCTTACGATTATGAGCAACTTTGGCGCGCGGAATTC 1669

RESULT 2  
STU249878 1641 bp mRNA linear PLN 04-JAN-2002  
LOCUS Solanum tuberosum mRNA for N-acetylglucosaminyltransferase I (GntI  
DEFINITION gene), clone A1.  
ACCESSION AJ249878  
VERSION AJ249878.1 GI:18076139  
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.  
SOURCE Solanum tuberosum  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (sites)  
Wendroth, I., Tjaden, J. and von Schaewen, A.  
Isolation and characterization of different plant  
N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional  
analyses in the Arabidopsis cgl mutant, and in potato and tobacco  
antisenese plants  
Unpublished  
2 (bases 1 to 1641)  
von Schaewen, A.  
Direct Submission  
Submitted (28-SEP-1999) by Schaewen A., Pflanzenphysiologie,  
Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,  
GERMANY

FEATURES  
Source  
1..1641  
/organism="Solanum tuberosum"  
/variety="Desiree"  
/db\_xref="taxon:4113"  
/clone="A1"  
/tissue\_type="leaf mesophyll"  
/note="sink"  
39..1379  
/gene="GntI"  
39..1379  
/EC\_number="2.4.1.101"  
/function="initiates complex-type glycans in the Golgi"  
/codon\_start=1  
/evidence=experimental  
/product="N-acetylglucosaminyltransferase I"  
/protein\_id="CAC80697.1"  
/db\_xref="GI:18076140"  
translation="MRGNKFCFLRLYLVLVVAALFIYIQLRFLATQSEYVDRLLAAIE  
AENHCTSRLLDKISQOGRVVALEOMKHODCEQLRALVQDLESKGIKKLIGD  
VOMPVAVVWACSRDYDLERTIKS ILKYOTSVASKYPLFISODGSPNDVKLALSYG  
QLTYMQLDYEPTHERPGLVAYYKIRHYKALDOLFHKNFSRVILLEDDEMTAA  
DFDYFAGATLLDRKSIKMAISWNDNGORQVQDDALYRSDFFPGLWMLSKSTW  
SELSPKWPAYDWDLRLKHNHRGROFIRPEVCTNFGHESSLGOFKQYLEP.LKL  
NDVQDWKSLSYLLLEDNYVKHFGLVKKAKPIHGADAVLKAFNIDGDVRIQYRDL  
DFEDIAROFIFIEWKDVPRAYKGIWVREFOTSRRVFLVSPDSLRLQIGVED"

BASE COUNT 485 a 302 c 377 g 477 t

Query Match 98.3%; Score 1641; DB 8; Length 1641;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 ctgagaaacctcgaaatcatttcgcatcttgagagatgagagggagaaagttttgt 74  
Db 1 CTGAGAAACCTTCGAATTCATTTCCGATTTGGCAGAGATGAGAGGGAACAAGTTTCT 60  
Qy 75 ttgatttaccggtacattctcgtcggtgctctcgccttcattctacatcacagatgcgc 134  
Db 61 TTGATTATTCGGTACCTTCTCGTCTGCTGCTCTCGCTTCATCTACATACAGATGCGGC 120  
Qy 135 ttttcgcacacagtcagaataatgtagaccgcttgcctgcgaattgaagcagaaaaac 194  
Db 121 TTTTTCGGCACAGTCAGAAATATGTAGACCGCTTGCCTGCTGCAATTTGAAGCAGAAATC 180  
Qy 195 attgtacaagtccagaccagattgcttattgcaaatagccagcagcaagaagtag 254  
Db 181 ATTGTACAAGTCAGACCAGATTGCTTATTGACAAGATTAGCCAGCAGCAAGAGTAG 240  
Qy 255 tagctttgaagaaacaaatgaagcatcaggaccaggagtgccggcaattaaaggctctg 314  
Db 241 TAGCTCTTGAAGAACAATGAAGCATCAGGACCGAGGTGCCGGCAATTTAAGGGCTCTTG 300  
Qy 315 ttcaagatcttgaagtaaggcataaaaaagttaatcgagatgtgcagatgcagtg 374  
Db 301 TTCAGGATCTTGAAGTAGGCGCATAAAAAGTTAATCGGAGATGTGCAGATGCCAGTGG 360  
Qy 375 cagctgtagttgttagcttcgctcagtcgactgactacctcgagagagactataaatcca 434  
Db 361 CAGCTGTAGTTGTTATGGCTTGCAGTCGTACTGACTACCTGGAGAGGACTATTAAATCCA 420  
Qy 435 tcttaaaataccaaacatctgttgcaataaaatactcttttcatatcccaggatggat 494  
Db 421 TCTTAAATACCAACATCTGTTGCATCAAAATATCCTCTTTTCATATCCCAGGATGGAT 480  
Qy 495 caaatcctgatgtaagaagcttgcctttgagctatggctcagctgacgtatatagcagcact 554  
Db 481 CAAATCCTGATGTAGAAAGCTTGTCTTTCAGCTATGGTCAGCTGACGTATATGACGACT 540  
Qy 555 tggattagaacctgtgcatactgaaagaccaggggaaactggtgcatactacaagatg 614  
Db 541 TGGATTATGAACCTGTGCATCTGAAAGACCAGGGGAACCTGGTTGCATACTACAAGATTG 600  
Qy 615 cagctcattacaagtggcgatgcatcagctgtttcacaaagcataatttagcgtgtta 674  
Db 601 CAGCTCATTAACAAGTGGGCATTTGGATCAGCTGTTTTCACAAGCATAAATTTAGCCGTGTTA 660  
Qy 675 tcatactagaagatgatattggaattgctgctgattttttgaactatttgaggtggag 734  
Db 661 TCATACTAGAAGATGATATGAAAATTTGCTGCTGATTTTTTTGACTATTTTGAAGCTGGAG 720  
Qy 735 ctactctttgcacagacaagtcgattatggctattcttcttggaalgacaatggag 794  
Db 721 CTACTCTTTTGACAGAGCAAGTCGATTATGGCTATATTTCTCTTGGATGACAATGGAC 780  
Qy 795 aaaggcagttcgtccagaatccctgatgctctttcacgcgcagactttttcctggtcttg 854  
Db 781 AAAGGCAGTTTCGTCGAAGATCTCTGATGCTCTTTTACCGCTCAGACTTTTTTCTCGTCTTG 840  
Qy 855 gatggatgctttcaaatcaactcgttcggaactatctccaaagtggccaaaggttact 914  
Db 841 GATGGATGCTTTCAAAATCAACTTGGTCGAACATATCTCCAAAGTGGCCAAAGGCTTACT 900  
Qy 915 ggatgaactgctcaagctgaaagaaaaatcacagaggtcgacaattatttcgcccagaag 974  
Db 901 GGGATGACTGGCTTAGCGTGAAAGAAAAATCACAGAGGTGCACAAATTTATTCGCCAGAG 960  
Qy 975 tttgcagaacgtacaatttttgggtgagcagttctgagttctggtgggcagttttttaagcagt 1034  
Db 961 TTTGCAGAACGTACAATTTTGGTGAGCAGTGTCTAGTTTGGGCAGTTTTTTTAAAGCAGT 1020  
Qy 1035 atcttgagccaataaagctaaaatgatgtccaggttgatgggaagtcaatggacctaagtt 1094

		/variety="Desiree" /db_xref="taxon:4113" /clone="A6" /tissue_type="tuber" /note="sink" 1..1298 /gene="GntI" /gene="GntI" /EC_number="2.4.1.101" /function="initiates complex-type glycans in the Golgi" /codon_start=3 /product="N-acetylglucosaminyltransferase I" /protein_id="CAC80698.1" /db_xref="GI:18076142" /translation="VAALAFIYIOMRLFATSEYVDRLEAAIAEDHCTQSOTRLLIDK ISQOGRVVALEQMKRQDQECROLRALVQDLESKIKKIGDVQMPVAHVWACSR ADYLEKTKISLKYOTSVASKYPLFI SQDGSNPQVRLALSLYDQLYMOHLDFEPVHT ERPELVAYYKIARHYKWLIDQLPHKHF SRVI ILEDDMEIAADFDFYFAGATLLDR DKSITMAISSMNDNGOROFVODPYALYRSDFFPGKGLWMLSKTWSLSPKPKPKAYLDDW LRLKXNHRGROFIRPEVCRTYNGEHSGLGQFKOYLEPIKLNDOVDMKSMDSYL LEDNYVHFGLVKKAPKIHGADAVLAKFNIDGVRIOYRDOLDFEDIARQFGIFEEM KDGVPRAAYKGI VVFRFOTSRVFLVAPDSLRQLGVEET"
gene	518 a 307 c 381 g 485 t	
CDS	BASE COUNT ORIGIN	
	Query Match 85.8%; Score 1432.6; DB 8; Length 1691; Best Local Similarity 96.7%; Pred. No. 0; Matches 1510; Conservative 0; Mismatches 34; Indels 17; Gaps 4;	
QY	98	gtggctgctgccttcacacagatgcgggcttttcgcgacacagtcagaatat 157
Db	3	GTGGCTGCTCTCGCTTCATCTACATACATGCGGCTTTTCGGCACACAGTCAGAATAT 62
QY	158	gtagaccgctgctgctgctcaattgaagcagaaatactgtcacagtcagaccagattg 217
Db	63	GTAGACCGCTTGCTGCTGCAATTGAAGCAGAAGATCATTTGTACAGTCAGACAGATTG 122
QY	218	cttattgacaagattagccagcagcaagagtagtagctcttgaagaacaaatgaag 277
Db	123	CTTATTGACAAGATTAGCCAGCAGCAAGGAGAGTAGTGGCTCTTGAAGAACAATGAAG 182
QY	278	catcagaccagagtgccggcaattaaggctctgttcaggactctgaagtaagggc 337
Db	183	CGTCAGCAGCAGGAGTCCCGCAANTTAAGGGCTCTTGTTCAGGATCTTGAAGTAGGGC 242
QY	338	ataaaaagttaacggagatgtgcagatgccagtgccagtgccgctgtagtgttatggctgc 397
Db	243	ATAAAAAAGTTAATCGGAGATGTGCAGATGCCAGTGCCGCTGTAGTTGTTATGGCTTC 302
QY	398	agtcgtactgactacactggaggagactataaatccatcttaaaataccaaacatctgt 457
Db	303	AGTCGTGCTGACTACCTGGAGAAGACTATTAAATCCCTCTTAAATACCAACATCTGTT 362
QY	458	gcataaaaatctctcttttcataccaggatggatcaaatccctgatgaagaagctt 517
Db	363	GCATCAAAATATCCTCTTTTCATATCCAGAGTGAATCAATCCTGATGTTAGAAAGCTT 422
QY	518	gctttgagctatggtcagctgacgtatatagcacacttgattatgaactgtgcatact 577
Db	423	GCTTTGAGCTTATGATCAGCTGACGTATATGCACACTTGATTTTGAACCTGTGCATCT 482
QY	578	gaagaccagggaactggttgcatatacaagattgcagctcattacaagtgggcattg 637
Db	483	GAAGACCAGGGGAACCTGGTTGCATACTACAAGATTGCAGTCAATTAACAAGTGGGATTG 542
QY	638	gatacgtgttcacaagaataatttttagccgtgttatcatatacagaataatagga 697
Db	543	GATCAGCTGTTTACAAAGCATAATTTTAGCCGTGTTATCATATAGTAAGATGATATGGAA 602
QY	698	attgctgctgattttttgactatttttgagctggagctactctcttgacagagacaag 757

Db	1021	ATCTTGAGCCATTAAAGCTAAATGATGTCCAGGTTCATTTGGAAGCTCAATGGACCTAAGTT 1080
QY	1095	accttttgagagacaactatgtgaacaacactttgacgacttggcttaaaaaagcctaagccca 1154
Db	1081	ACCTTTTGAGGACAACTATGTGAACAACTTTGGCGACTTTGGTTAAAAAGGCTAAGCCCA 1140
QY	1155	tcacagagactgatgctgttttgaagcatttaacatagatgggtgatgtgcgtattcagt 1214
Db	1141	TCCACGGACTCATGCTGTTTGAAGAGCATTTAAATAGATGGTGCATGTGGGTATTTCAGT 1200
QY	1215	acagagaccaactagactttgaagatatcgtcgcagagtttggcatttttgaacaatgga 1274
Db	1201	ACAGACACCAACTAGACTTTGAAGATAATCGCTCGACAGTTTGGCATTTTTCGACGAATGGA 1260
QY	1275	aggatggtgtaccacgggcagcagataaaggagtagtagttttccgggtttcaaacatcta 1334
Db	1261	AGGATGGTGTACCACGGGCAGCATATAAAGGATAGTAGTTTTCGGGTTTCAAAACATCTA 1320
QY	1335	gacgtgttctctgtttccctcgtattcttcgcacaacttggagttgaagatacttagc 1394
Db	1321	GACGTGTGTTCCCTGTGTTTCCCTCGATTCCTTCGACAACCTTGGAGTTCGAAGATCTTAGC 1380
QY	1395	gaagatatgtggagcctgagcaacaatttagacctatttggtaggatacalttgaag 1454
Db	1381	GAAGATATGATTGAGGCCCTGAGCAACAATTTAGACTTATTTGGTAGGATACATTTGAAAG 1440
QY	1455	agctgacagaaaagtatgactaccagtagctacatgcaacatttttaattgaatggag 1514
Db	1441	AGCTGACACGAAAGTATGACTTACCAGTAGCTACATGCAACATTTTAATGTTAATGGAG 1500
QY	1515	gaacccactgcttattgtytggatggatgaatcatcacacactctattattcaagtta 1574
Db	1501	GAACCCACTGCTTATTGTTGGATGGATGAATCATCACACATCTATTATTTCATCAAGTTTA 1560
QY	1575	caacataaagagaaaagtgtgccctataaaacaaaatttttgttctcaagaaggaacg 1634
Db	1561	CAACATAAAGAGAAATGTTGCCCTATAAAAACAAATTTTGTGTTCTTAAGAAGGAACG 1620
QY	1635	ttacgattatgacaacttgg 1655
Db	1621	TTACGATTATGAGCAACTTTG 1641
RESULT	3	
STU249879		
LOCUS	1691 bp mRNA linear	PLN 04-JAN-2002
DEFINITION	Solanum tuberosum partial mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A6.	
ACCESSION	AJ249879	
VERSION	GI:18076141	
KEYWORDS	GntI gene; N-acetylglucosaminyltransferase I.	
SOURCE	potato.	
ORGANISM	Solanum tuberosum	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.	
AUTHORS	1 (sites)	
TITLE	Wenderoth, I., Tjaden, J. and von Schaewen, A.	
	Isolation and characterization of different plant N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 1691)	
TITLE	von Schaewen, A.	
JOURNAL	Direct Submission	
	Submitted (28-SEP-1999) by Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY	
FEATURES	Location/Qualifiers	
SOURCE	1..1691	
	/cgcgcgcgc="Solanum tuberosum"	

Db	603	ATTGCTGCTGATTTTTTTTTCACATATTTTTCAGGCTGGAGCTACTCTTCTTGACAGAGATAAG	662
Qy	758	tcgattatggctattcttcttggaaatgacaatgagcaaaaggcagttcgtccgaagatccct	817
Db	663	TCGATTATGCTATTTCTCTTTCGGAATGACAATGACAAAGCGCAGTTGCTGCAAGATCCT	722
Qy	818	gatgctcttaccgtcagacttttttctgctctgctgctgctgcttccaaatcaact	877
Db	723	TATGCTCTTTACCGCTCAGACATTTTTTCTGCTGCTGATGATGCTTTCAAAATCAACT	782
Qy	878	tggtccgaactatctccaaatgyccaaggcttactgggagctggtgactgagcgtgaaa	937
Db	783	TGTTCCGAACATATCTCAAAGTGGCAAGGCTTACTGGGATGACTGGCTAAGGCTCAA	842
Qy	938	gaaaatcacagagtcgacaatttatctgccagagtttgcagaacgtacaaattttggt	997
Db	843	GAAATCACAGAGTGCACAATTTATTCGCCCAGAAAGTTTCAGAACATACAAATTTTGGT	902
Qy	998	gagcatggttctagtttggggcagttttttaagcagttatcttgagccaaattaagctaaat	1057
Db	903	GAGCATGTTCTAGTTTGGGGCAGTTTTTTTAAGCAGTATCTTTGAGCCAATTAAGCTAAAT	962
Qy	1058	gattccaggttattggaagtcaatggacctaaagttaacctttttggggacaactatgtg	1117
Db	963	GATCTCCAGGTTGATTGGAAGTCAATGGACCTTAAGTTACTCTTTTGAGGACAACTATGTG	1022
Qy	1118	aaacactttggcacttgggttaaaaggctaaagcccatccacggagctgctgtcttttg	1177
Db	1023	AAACACTTTGGCGACTTGGTTAAAAAGGCTTAAGCCCATCCAGCGTCTGATGCTGTTTG	1082
Qy	1178	aaagcatttaacatagatggtgctgctattcagttacagtcagagacaactgaactttgaa	1237
Db	1083	AAAGCATTTAACAATAGATGTGATGTGCGTATTTCAGTACAGAGCACTACATGCTTGA	1142
Qy	1238	gatacgtctgcacagtttggcatttttgaagaatggaagatgggtgtaccacgggcagca	1297
Db	1143	GATATCGCTCGCAGTTTGGCATTTTGAAGAAATGGAAGGATGGTGTACCACGGGCAGCA	1202
Qy	1298	tataaaggatagtagtttccgggttccaaacatctagacgtgtgtctctgtttccctc	1357
Db	1203	TATAAAGGGATAGTAGTTTTTCCGGTTTCAACATCTTAGACGTGTCTCTGTGTGCCCT	1262
Qy	1358	gattctctgcacaaacttgagttggaagatacttagcgaagatatga-ttgagcctgag	1416
Db	1263	GATTCCTCTCCACAACCTGGAGTTGAAGATACCTAGCGAAGATATGATTTGGTGCCTGAG	1322
Qy	1417	caacaatttagactatttggtaagatacatatttgaagagctgacacgaaaagtatgact	1476
Db	1323	CAACACTTTAGACTATTTTGGTAGGATACATTTGAAGGAGCTGACACGAAAGCATGACT	1382
Qy	1477	accagtagctacatgcacattttaaqttaatggaagggaaccactgcttattgttga	1536
Db	1383	ACCAGTAGCTACATGCACATTTTAAATGTTAATGGAAGGAACCTCACTGCTACTGTTGA	1442
Qy	1537	atggatgaatc-atcacacatctctattattcaag-tttcaacaataaagagaataatgt	1594
Db	1443	ATGGATGAATCAATCACCACAGCCTATTATTCAAGTTTTCACAAACATAAAGAGG----	1496
Qy	1595	tgccctataaaacaaaattttttgttcttaagaaggacgttacgattatgagcaacttt	1654
Db	1497	-----GACACATGTTTTTTTCTTCTAAGAGGAGCGTTACGATTATGAGCGACTTT	1548
Qy	1655	g 1655	
Db	1549	G 1549	
RESULT	4		
STU249880			
LOCUS			
DEFINITION	STU249880	1499 bp	mRNA linear
ACCESSION	STU249880		PLN 04-JAN-2002

VERSION	AJ249880.1	GI:18076143	
KEYWORDS	GntI gene; N-acetylglucosaminyltransferase I.		
SOURCE	potato.		
ORGANISM	Solanum tuberosum		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.		
AUTHORS	1 (sites)		
TITLE	Wenderoth, I., Tjaden, J. and von Schaewen, A.		
JOURNAL	Isolation and characterization of different plant		
REFERENCE	N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 1499)		
JOURNAL	von Schaewen, A.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie,		
TITLE	Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,		
JOURNAL	GERMANY		
FEATURES	Location/Qualifiers		
source	1..1499		
	/organism="Solanum tuberosum"		
	/variety="Desiree"		
	/db_xref="taxon:4113"		
	/clone="A8"		
	/tissue_type="tuber"		
	/note="sink"		
gene	37..1158		
	/gene="GntI"		
CDS	37..1158		
	/gene="GntI"		
	/EC_number="2.4.1.101"		
	/function="initiates complex-type glycans in the Golgi"		
	/codon_start=1		
	/product="N-acetylglucosaminyltransferase I"		
	/protein_id="CAC80899.1"		
	/db_xref="GI:18076143"		
	/translation="MKRQDECRQLRALVODLESKIKKILGDVQMPVAVVVMACSR		
	ADYLERIKSLKYQTSVASKYPLFI SQGSDPNVRKLALSLYDQLYMHQLDFEPVHT		
	ERPELVAYVKIARHYKALDQLFHKNFSRVILLEDMEIAADFDFEAGATLLDR		
	DKSTMATSSWNDGROFVODPALYRSDFPGLGWLMSKSTWSELSKPKAYWDDW		
	LRKENHGRQFIHPEVCRTYNEGHSGISGFOFKOYLRPIKLNDVQVDMKSMDSL		
	LEDNVYKFDGLVKKAKPIHGADAVLKAFNIDGDRIOYRDQLDFAIARQFCIFEEW		
	KDGVPRAYKGIIVFRFQTSRRRFLVAPDSLRQLGVEDT"		
BASE COUNT	452 a	266 c	339 g
ORIGIN			442 t
Query Match	81.2%	Score 1356;	DB 8; Length 1499;
Best Local Similarity	98.5%;	Pred. No. 0;	
Matches 1379;	Conservative	0; Mismatches	20; Indels 1; Gaps 1;
Qy	257	gctcttgaagaacaaatgaagcatcaggaccaggagtgccggcaataaaggctctgtt	316
Db	22	GTTCCTCGGTAAACAAATGAAGCGTCAGGACGAGTGCCGGCAATTAAAGGCTCTTGT	81
Qy	317	caggatcttgaagtaagggcataaaggtaataatcgagatgtgcagatgccagtggca	376
Db	82	CAGGATCTTGAAGTAAGGGCATAAAAAGTTAATCGGAGATGTGCAGATGCCAGTGCA	141
Qy	377	gctgtagttgttattggcttgctgactgactacacctggagaggactattaaatccatc	436
Db	142	CGTGTAGTTGTTATGGCTTGCGATCGTCTACTACTCTGGAGAGGACTATTAAATCCATC	201
Qy	437	ttaaataaccacaaatctgttgcatacaaaatatcctctttcatatccagatggatca	496
Db	202	TTAAATAACCAACATCTGTTGCATCAAAATATCCTCTTTTCATATCCAGATGGATCA	261
Qy	497	aatcctgatgaagaagcttctgttgactatgctgacgtgacgtatataatcagacttg	556
Db	262	AATCTGTATGAAGAAAGCTTCTTTGAGCTATGATGATGACGTGACGTATATGACGACTTG	321

Qy	557	gattatgaacctgtgcatactactaaagaccaggggaactggttgctatactactacaagattgca	616
Db	322	GATTTTGAACTGTGCATACTCAAGAGACCAGGGAACGGTTGCATACTACAAGATTGCA	381
Qy	617	cgtcattacaagtgggcattgatgcagctggtttccaaagcataatttttagccgtgttatc	676
Db	382	CGTCATTACAAGTGGGCATTGGATCAGCTGTTTCAACAAGCATATTTAGCCGTCTTATC	441
Qy	677	atactagaagatgatatgaattgctgctgattttttttagctatttttgaggctgagct	736
Db	442	ATACTAGAAGATGATATGAANAATGCTGCTGATTTTGTGACTATTTTGAGCGCTGGAGCT	501
Qy	737	actctcttgacagagacaagtcgattatggtctattctcttggaaatgacaatggacaa	796
Db	502	ACTCTTCTTGACAGACACAAGTCGATTATGGCTATTTCTTGTGAATGACAATGGACAA	561
Qy	797	aggcaatcgctcaagatcctcatctctttaccgctcagacttttttccctgctctgga	856
Db	562	AGCAGTTCGTCAGATCCTGATGCTCTTTACCGCTCAGACTTTTTCCTGGTCTTGA	621
Qy	857	tggatgctttcaaaatcaacttggctcgaaactatctccaaaatggccaaaaggcttactgg	916
Db	622	TGCATGCTTTCAAAATCAACTTGGTCCGAACATATCTCCAAGTGGCCAAAGGCTTACTGG	681
Qy	917	gatgactggcctaaggctgaagaaaaatacacagagtcgaaggtcgaatttatcgccaaagt	976
Db	682	GATGACTGGCTAAGGCTGAAGAAATAACAGAGGTCGACAATTTATTCACCCAGAAGTT	741
Qy	977	tcgagaacgtacaaatttggtagcagtctcttagtttggggcagtttttttaagcagtat	1036
Db	742	TGCAGAACGTACAAATTTTGGTGAGCATGCTTCTAGTTTGGGCAATTTTAAAGCAGTAT	801
Qy	1037	cttgagccaaattaagctaaatgatgtccaggttgattggaagtcgaatgacctaagttaac	1096
Db	802	CTTGAGCCCAATTAAAGTAATAATGATGCCAGGTTGATTGGAAGTCATATGACCTAAGTTAC	861
Qy	1097	cttttgaggacaactatgtgaacaactttggcgacttgggtttaaagaagcctaagccatc	1156
Db	862	CTTTTGAGGACAACATATGTGAACACTTTGGCGACTTGGTTAAAAAGGCTAAGCCCATC	921
Qy	1157	cacggagctgactgctgttttgaagcatttaacatagatgggtgagtgcgattcaagtaac	1216
Db	922	CACGGAGCTGATGCTGTTTGAAGCATTTAACATAGATGGTGATGGCTGATTCAGTAC	981
Qy	1217	agagacaaactagacttgaagatatcgctcgacagtttggcatttttgaagaatggaag	1276
Db	982	AGAGACCAACTAGACTTTGAAGCTATCGCTCGACAGTTTGGCAATTTTGAAGAAATGGAAG	1041
Qy	1277	gatggtgtaccacgggcagcataataaaggatagtagttttccgggtttcaaacatctaga	1336
Db	1042	GATGGTGTACCACGGGCAGCATATAAAGGATAGTAGTTTTCGGGTTTCAAAACATCTAGA	1101
Qy	1337	cgtgtgtctctgttttccctgattctcttcgacaacttggagttgaagatacttaagca	1396
Db	1102	CGTGTGTTCTGTTGCCCTGATTCCTTCGACAACCTTGGAGTTGAAGATACATTAGCGA	1161
Qy	1397	agatatgattggagcctgagcaacaatttagacttatttggtaggatacatatttgaagag	1456
Db	1162	AGATATGATTGGAGCCTGAGCAACAATTTAGACTTATTTGGTAGGATACATTGGAAAGAG	1221
Qy	1457	ctgacacgaaaaatgactaccagtagctacatgcaacaatttttaattgaatggaagga	1516
Db	1222	CTGACACGAAAAGTATGACTACCAGTAGCTACATGTCATGCAACATTTTAATGTAAATGGAAGGA	1281
Qy	1517	accactgcttattgttggaaatggaatcatcatccacacatcctattattcaag-cttac	1575
Db	1282	ACCACCTGCTTATTGTGGAAATGGATGAATCATCACCACATCCTATATTCAAGTTTAC	1341
Qy	1576	aaacataaagagaaatggtgcctataaaaaacaaattttttgttcttaagagggaacgt	1635
Db	1342	AAACATTAAGAGGAAAATGTAGCCCTATAAACACAATGTTTTTGTCTTAAGAAGGAACGT	1401
Qy	1636	tacgattatgacaaactttg 1655	

NTY16832

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

5'UTR

CDS

3'UTR

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Mismatches

Conservative

Indels

Gaps

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy



NDVOYDMKSMDSLVSLLDNYVKHFGDLVKKAKPIHCADAVLKAFNIDGVRIOYBDOL  
DFENIARFGFEEHKDGVPRAAAYKGIYVFRYQTSRRVFLVGHDSLQQLGIEDT\*  
BASE COUNT 506 a 326 c 380 g 496 t  
ORIGIN

Query Match 75.2%; Score 1255; DB 8; Length 1708;  
Best Local Similarity 89.1%; Pred. No. 0;  
Matches 1412; Conservative 0; Mismatches 140; Indels 33; Gaps 4;

QY 52 gatgagagggaacaagttttgcttgatttacggtacccttctcogtggctctctcgc 111  
|||||  
Db 112 GATGAGAGGGAACAAGTTTGTGCTGATTCCGGTACCTCCTCATCTTGGCTCTGTCG 171  
|||||  
QY 112 cttcatctacatacagatcgcgcttttcgcgacacagtcagaataatgtagacgcgcttgc 171  
|||||  
Db 172 CTTTCATCTACACAGATCGCGCTTTTTCGCACACAGTCAGAAATATGCAGATCGCCTTGC 231  
|||||  
QY 172 tgcgcaattgaagcagaaaaatcattgtacaagtcagaccagatgcttattgacaagat 231  
|||||  
Db 232 TCGTGCAAATGAAGCAGAAAAATCATGTACAGCCAGACAGATTGCTTATTGACCAGAT 291  
|||||  
QY 232 tagccgcagcgaaggaagtagtagctcttgaagacaaaatgaagcatcagagaccagga 291  
|||||  
Db 292 TAGCCTGCACGACGAAGGAATAGTTGCTCTTGAAGAACAAATGAAGCGTCAGCACCAGGA 351  
|||||  
QY 292 gtccgcgaattaaaggtcttctgttcaggatcttgaagaatgaagggcataaaaaagttaat 351  
|||||  
Db 352 GTCCCCACAAATTAAGGGCTCTTGTTCAGGATCTTGAAGTAAGGGCATAAAAAGTTGAT 411  
|||||  
QY 352 cggagatgtcgagatgccagtgccagctgtagtgtgtatggtctgcagtcgactgaacta 411  
|||||  
Db 412 CGGAATGTACAGATGCCAGTGGCTGCTGTAGTTGTATTGCTTGAATCGGCTGAATTA 471  
|||||  
QY 412 cctggagaggactattaaatccatctttaaataaccaaacatctgttgcatacaaaatctcc 471  
|||||  
Db 472 CCTGGAAAAGACTATTAAATCCATCTTAAATACCAAAATATCTGTTCGCTCAAAATATCC 531  
|||||  
QY 472 tcttttcatacccagatggatacaaatcctgatgtaagaagccttgcttgaagctatgg 531  
|||||  
Db 532 TCTTTTCATATCCCAGATGGATACATCTCTGATGTCAGGAAGCTTGGCTTGAAGCTATGA 591  
|||||  
QY 532 tcagctgacgtatgacgacttgattatgaacctgtgcatactgaaagaccagggga 591  
|||||  
Db 592 TCAGCTGACGTATATGCACACTTGGATTTGCAACTGTGCATACCTGAAAGACCAGGGGA 651  
|||||  
QY 592 actggtgcatactacaagatgcagctcattacaagtgggcattggatcagctgttca 651  
|||||  
Db 652 GCTGATTGCATACTACAAAATTCACAGTGCATTTACAAGTGGCATTTGCATCAGCTGTTTTA 711  
|||||  
QY 652 caagcataattttagccgtgttatactatacagaagatgatggaattgctgctgattt 711  
|||||  
Db 712 CAAGCATAATTTTACCGTGTTTATCATACTAGAAGATGATATGGAATTTGCCCTTGATTTT 771  
|||||  
QY 712 ttttgactattttgaggctggagctactcttctgacagagacaagtcgattatgctat 771  
|||||  
Db 772 TTTTGGACTTTTGGAGCTGGAGCTACTCTCTTGACAGACACAGTCGATTTATGGCTAT 831  
|||||  
QY 772 tctcttggaaatgcaatggcaagcagtcgttcccaagatcctgtagctctttaccg 831  
|||||  
Db 832 TTTCTCTTGGAAATGACAATGGACAAATGCAGTTTGTCTCCAAAGATCCTTATGCTTTTACC 891  
|||||  
QY 832 ctccagcttttctgctgttggatgcttccaaatcaacttggtccgaactatc 891  
|||||  
Db 892 CTCAGATTTTTTCCCGTCTTGGATGGATGCTTTCAAATCTACTTGGGACGAATATTC 951  
|||||  
QY 892 tccaaatggccaaaggcttactggatgtagctggctgaagcgtgaaagaaatcacagag 951  
|||||  
Db 952 TCCAAAGTGGCCAAAGGCTTACTGGGACGACTGGCTTAAGACTCAAGAGAAATCACAGAG 1011  
|||||  
QY 952 tcgacaattttatcccccagaagtttcgagaacgctacaattttgtagagcgtggttctag 1011  
|||||  
Db 1012 TCGACAATTTATTCCGCCAGAAGTTTGACAGACATATAATTTTGTGTGAGCATGGTCTTAG 1071  
|||||

QY 1012 tttggggcagcttttttaagcagtatctcttgagccaatttaagctaataatgatgtcccsggttga 1071  
|||||  
Db 1072 TTTGGGCGAGTTTTTCAAGCAGTATCTTGAGCCAATTAATTAATGATGTCCAGGTTGA 1131  
|||||  
QY 1072 ttgaaagtcaatgagcctcaagtacccttttggagagacaactatgtgaaacactttgcga 1131  
|||||  
Db 1132 TTGAAGTCAATGAGCCTTAGTTACCTTTTGGAGGACAATTACGTGAACACTTTTGGTGA 1191  
|||||  
QY 1132 cttgggttaaaaggcctaagcccccacacgagcgtgtagctgttttgaagcatttaacat 1191  
|||||  
Db 1192 CTTGGTTAAAAAGCTAAAGCCCATCCATGAGCTGATGCTCTCTTGAAGCATTTAAACAT 1251  
|||||  
QY 1192 agatggtgtagtcgctattcagtcacagagaccacacagacttttgaagatatcgctcga 1251  
|||||  
Db 1252 AGATGCTGATGTGCGTATTTCAGTACAGAGATCAACTAGACTTTTGAAATATCGCACGGCA 1311  
|||||  
QY 1252 gtttggcatttttgaagaatggaagatggtgtaccacggcagcatataaaaggatagt 1311  
|||||  
Db 1312 ATTTGGCATTTTGAAGAATGGAAGGATGGTGTACCACGTGCACGATATAAAGGAATAGT 1371  
|||||  
QY 1312 agtttccgggttccaacatctagacgtgtgttcccttgtttccctctgattctcttcgaca 1371  
|||||  
Db 1372 AGTTTTCGGTACCAAAACGTCAGACGCTGATTCCTTGTGTGGCCATGATTCGCTTCAACA 1431  
|||||  
QY 1372 acttgaggttgaagatactttagcgaagatgatt---ggagcctgagcacaacatttga 1428  
|||||  
Db 1432 ACTCGGAATTTGAAGTACTTTAACAAGATATGATTCGAGAGCCCGGGCAAAATTTTGA 1491  
|||||  
QY 1429 cttatttggtagcatcatttgaagagctgcacagaaagatgactgactaccagtagctac 1488  
|||||  
Db 1492 CTTATTGGGTAGGATGCATCGAGCTGACACTAAACCATGATTTT-----ACCAAGTTAC 1544  
|||||  
QY 1489 atgcaacattttaatgttaatggaaggaacccactgct-----t 1527  
|||||  
Db 1545 ATACACGTTTTTAATGTTATACGGAGGAGCTACTGTTCTAGTGTTCGAAGGATATCGGC 1604  
|||||  
QY 1528 attgttggaaatgagatgaatcatcacacatcctatta--ttcaagttttacaacataaag 1585  
|||||  
Db 1605 TTTCTTAGTATTGGATGAATCATCAACACAACCTATTATTATTAAAGTGTTCAGAACATAAG 1664  
|||||  
QY 1586 aggaatgttgcctcataaaaaaa 1610  
|||||  
Db 1665 AGGAATGTAGCCTGTAAAGACTA 1689  
|||||  
RESULT 7  
A95055  
LOCUS A95055 1737 bp DNA linear PAT 26-JAN-2000  
DEFINITION Sequence 3 from Patent WO9929879.  
ACCESSION A95055  
VERSION A95055.1 GI:6779208  
KEYWORDS common tobacco.  
SOURCE  
ORGANISM  
Nicotiana tabacum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Nicotiana.  
Von,S.A.  
VEGETABLE GntI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH  
A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GntI)  
ACTIVITY  
Patent: WO 9929879-A 3 17-JUN-1999;  
VON SCHAEWEN ANTJE (DE)  
JOURNAL Location/Qualifiers  
FEATURES  
source  
1. 1737  
/organism="Nicotiana tabacum"  
/strain="SAMSUN NN"  
/db\_xref="taxon:4097"  
/cell\_type="BLATTZELLEN"  
/tissue\_type="MESOPHYLL"  
/clone\_lib="LAMBDA ZAP II (ECORI)"



```
/clone_lib="(B) CLON(E): GNTI-A9(T)"
/dev_stage="M: 'SINK' ORGAN"
15..126
127..1467
/genes="CGL"
127..1467
/genes="CGL"
/EC_number="2.4.1.101"
/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.
GLYKOPROTEINEN"
/notes="ERSTE GNTI-SEQUENZ AUS TABAK (UNPUBLIZIERT)"
/codon_start=1
/evidence=experimental
/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I"
/protein_id="CAB70464.1"
/db_xref="GI:6779209"
/translation="MRGNKFCDFRYLLILAAVFIYQMRLFATQSEYADRLAAIE
AENHCTSTRLLIDIGISLQOEKQVLEQMKRQDOECQLRALVDLSEKGIKLIGN
VQMPAAVVVWACNRADYLEKTIKSLKQISVASKYPLFIQDSGSHDPVKLALSYD
QUTYMHQDLDFEPVHTERPEGLIAYYKIHRYKQWALDOLFYKHNFSSRVILLEDMEIAP
DFFDFEAGATLLDRKDSIMATSSWNDNGOMQFVODPYALYRSDFPGLGWLMSKSTW
DELSKPWPAYWDDMLRLKENHGRGOFIRPEVCRTYNEGEHGSLSLQGFQYLEPIKL
NDVOVDMKSMDSLYLLENYKHHKRGDLYKKAKPIHGADAVLKAFNIDGVDRIQVRDL
DPENIAROFGIPEENKGVPRAAKYGVVFRYQTSRRVFLVGHDSLQQLGIET"
<154..>5213
/genes="CGL"
/standard_name="MEMBRANANKER EINES TYP II GOLGI-PROTEINS"
/function="MEMBRANANKER (AMINO ACID 10-29)"
/notes="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN
IDENTIFIZIERT"
/codon_start=1
/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"
/protein_id="CAB70465.1"
/db_xref="GI:6779210"
/translation="FRVLLILAAVAFIVTQMLRF"
1468..1723
510 a 337 c 390 g 500 t
3'UTR
ORIGIN
Query Match 75.2% Score 1255; DB 6; Length 1737;
Best Local Similarity 89.1% Pred. No. 0;
Matches 1412; Conservative 0; Mismatches 140; Indels 33; Gaps 4;
QY 52 gatgagagggaacaagtttctgttgattacggttacctctcgtcggtgctctcgc 111
DB 126 gatgagagggaacaagtttctgttgattacggttacctctcgtcggtgctctcgc 111
QY 112 ctctcatclacaacagatgcggcttttcgcgacacagtcagaatatgacacgcttgc 171
DB 186 ctctcatclacaacagatgcggcttttcgcgacacagtcagaatatgacacgcttgc 171
QY 172 tctgcaattgaagcagaaatcattgacaaagtcagaccagattgcttattgacaagt 231
DB 246 tctgcaattgaagcagaaatcattgacaaagtcagaccagattgcttattgacaagt 231
QY 232 tagccagcagcagaagagatgactctcttgaaagacaaatgaagcatcaggacaagga 291
DB 306 tagccagcagcagaagagatgactctcttgaaagacaaatgaagcatcaggacaagga 291
QY 292 gtgcgcgaattgaaggctctgttcaggatcttgaaagtgaaggcataaaaaagtaatt 351
DB 366 gtgcgcgaattgaaggctctgttcaggatcttgaaagtgaaggcataaaaaagtaatt 351
QY 352 cggagatgtcaatcaccagtcagctgactgttattgcttgcagtcgactgacta 411
DB 426 cggaaatgtacacatgccatgctgctgtgattgttattgcttgcagtcgactgacta 411
QY 412 cctggagagactataatccatcttaaaataccaaacatctgttgcatcaaaatacc 471
DB 486 cctggaaacactatttaaatccatcttaaaataccaaacatctgttgcatcaaaatacc 471
QY 1528 attgttgaagtgaatcatcaccacatcctatta--ttcaagtttacaacaataag 1585
```

```
QY 472 tctttcataatccagatgatgatacaatcctgatgttaagaaagcttgccttgagctatgg 531
DB 546 TCCTTTTCATATCCAGGATGGATGCATCCTGATGTCAGGAAGCTTGTGAGCTATGA 605
QY 532 tcagctgacgtatatatcagcacttggattatgaacctgtgcatactactgaagaccaggga 591
DB 606 TCAGCTGCAGCTATATCAGCAGCTTGGATTTGAACTGTGCATCTGAAAGACACAGGGA 665
QY 592 actgggtgcataactacaagatgcagctcattacaaagtggcgatggatcagctgttca 651
DB 666 GCTGATTGCTACTACAAAATTCAGCTGATACAAAGTGGGATTCGATTCAGCTGTGTTA 725
QY 652 caagcataatttagcgtgttatcatactactagaagatgatggaaatgctgctgattt 711
DB 726 CAAGCATAAATTTAGCCGTGTTATCATACTAGAAGATGATATGGAATTTGCCCTGATTT 785
QY 712 ttttgaactattttagcgtgttgagctactcttctgaagagacaagtcgattatggctat 771
DB 786 TTTTGACTTTTTTGGAGCTGGAGCTACTCTTCTGACAGAGACAAGTCGATTTATGCTAT 845
QY 772 ttctcttggaaatgacaatggacaaagcagttcgtccaagatcctgatgctctttaccg 831
DB 846 TTCCTCTTGGAAATGACAATGGACAATCGAGTTTGTCCAAGATCCTTATGCTCTTTACCG 905
QY 832 ctacgaacttttctcgtgcttggatgagatgcttccaatacaactggtccgaactatc 891
DB 906 CTCAGATTTTTTCCCGGCTTTGGATGGATGCTTTCAAAATCTACTTGGAGCAATTTATC 965
QY 892 tccaaagtggccaaaggcttactgggatgactggcgaagctgaagctgaagaaatcacaggg 951
DB 966 TCCAAAGTGGCCAAAGGCTTACTGGGACGACTGGCTAAGACTCAAGAGAAATCACAGAGG 1025
QY 952 tcgacaattttatccgcccaagtttgcagacgtacaaattttggtagcatggtctctag 1011
DB 1026 TCGACAATTTATTCGCCCAGAAAGTTTGCAGAACATATAAATTTGCTGAGCATGTTCTTAG 1085
QY 1012 ttggggcagttttttaaagcagatcttgaagcacaatgaagtaaatgatgtccaggttga 1071
DB 1086 TTTGGGCGAGTTTTTCAACGAGTATCTTTGAGCCAATTAACATAATGATGTCCAGGTTGA 1145
QY 1072 ttgaaagtcgaatggacctaagttacccttttggaggaacactatgtgaacactttgcga 1131
DB 1146 TTGAAAGTCAATGGACCTTAGTTACCTTTTGGAGGACAAATTTACCTGAAACACTTTTGTGA 1205
QY 1132 ctgtgttaaaagcctaagccatccacgcagcgtgatgctgttttgaagcatttaacat 1191
DB 1206 CTGTGTTAAAGAGGCTAAGCCCATCCATGGAGCTGATGCTGTCTTTGAAAGCATTTAAACAT 1265
QY 1192 agatgggtgatgtgcgtatttcagtcacagacacacactagactttgaagatatcgtcga 1251
DB 1266 AGATGGTGATGTGGGTATTTCAGTACAGAGATCAACTAGACTTTGAAATATATCGACGGCA 1325
QY 1252 gtttggcattttgaagaatggaaggtggtgtaccacggcgagcatataaaggatagt 1311
DB 1326 ATTTGGCATTTTTTGAAGAAATGGAAGGATGGTGTACCACGTCGACGATATAAAGGAATAGT 1385
QY 1312 agtttccggtttccaacaatctagacgtgttctcgttttccctgattctcttcgaca 1371
DB 1386 ASGTTTTCCGGTACCAAAAGCTCCAGAGCTGATTTCCCTTTGTCCTTTGAGCCATGATTCGCTTCAACA 1445
QY 1372 acttggagttgaaga tacttagcgaagatatgatt---ggagcctgagcaacaatttga 1428
DB 1446 ACTCGGAATTTGAAGATACTTAACAAAGATATGATTGCAGAGGCCCGCGCAAAATTTTGA 1505
QY 1429 ctattttgtagatatacatttgaagagctgcacagaaagtatgactaccagtagctac 1488
DB 1506 CTATTGGGTAGATGCATCGAGCTGACACTAAACCATGATTTT-----ACCAGTTAC 1558
QY 1489 atgcaacattttaattgttaattggaaggaacccactgct-----t 1527
DB 1559 ATACAACGTTTTTAATGTTATACGGAGGAGCTCAGTGTTCAGTGTGTAAGGGATATCGGC 1618
QY 1528 attgttgaagtgaatcatcaccacatcctatta--ttcaagtttacaacaataag 1585
```



Db	1619	TTCTTAGTATTGGATGAATCATCAACCAACCTATTATTTTAAGTGTTCAGAACATAAAG	1678
Qy	1586	aggaatgttgcctataaaaaacaa	1610
Db	1679	AGGAAATGACGCTGTAAAGACTA	1703
RESULT	8		
NTA249882		1836 bp	linear
LOCUS		Nicotiana tabacum mRNA for N-acetylglucosaminoltransferase I (GntI gene), clone A4.	PLN 04-JAN-2002
DEFINITION			
ACCESSION		AJ249882	
VERSION		AJ249882.1	GI:18076145
KEYWORDS		GntI gene; N-acetylglucosaminoltransferase I.	
SOURCE		common tobacco.	
ORGANISM		Nicotiana tabacum	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.	
AUTHORS		Wenderoth, I., Tjaden, J. and von Schaewen, A.	
TITLE		Isolation and characterization of different plant N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 1836)	
AUTHORS		von Schaewen, A.	
TITLE		Direct Submission	
JOURNAL		Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY	
FEATURES		Location/Qualifiers	
source		1..1836	
		/organism="Nicotiana tabacum"	
		/variety="Samsun NN"	
		/db_xref="taxon:4097"	
		/clone="A4"	
		/tissue_type="leaf mesophyll"	
		/note="sink"	
gene		27..1367	
		/gene="GntI"	
CDS		27..1367	
		/gene="GntI"	
		/EC_number="2.4.1.101"	
		/function="initiates complex-type glycans in the Golgi"	
		/codon_start=1	
		/product="N-acetylglucosaminoltransferase I"	
		/protein_id="CAC80701.1"	
		/db_xref="GI:18076146"	
		/translation="MRGYKPCDDFRYLILAAVAFTYIQMLRFATSEYADRLAAAEI AENHCTSLRLIDQISQOGRIVALEQMKRQDQECRLALVQDLESKGTKKLIG VOMPAVVMACNRADYLEKTSILKYOISVAPKPLFIQDGSHPDVRKIALSYD QLTYMOHLDVFPVHTBPGEELIAYKIARHYKWLDDOLFVKHFSRVILLEDMEIAT DFDPEAGATLLDRKSTMAISSNDQMGQFVQDPYALYRSDFPGLCWLMSKSTW DELSFKPKAYKWDLLKLNHRGRGFIPEVCRSYNFGHGSSUGGFFPKYLEPIKL NDVQVDMKSNLYLLEDNYVHFGLVKKKAPIHGADAVLKAFNIDGVRVRIQYRDQL DFSDIARQGIIEEWDGVPRAAYGVFVRYQTSRRLVFLGPDLSLQQLGNEDT"	
BASE COUNT	536 a	354 c	403 g
ORIGIN			543 t
Query Match	74.7%	Score 1246.8	DB 8; Length 1836;
Best Local Similarity	91.1%	Pred. No. 0;	
Matches 1359; Conservative	0	Mismatches 127;	Indels 6; Gaps 3;
QY	52	gatgagaggaacaaagtttgcgtttgatattacggtacccttcgtcgtagctctcgc	111
Db	26	GATGAGAGGGTACAAGTTTTCGTGTGATTTCCCGTACCTCCTCATCTTGGCTGCTGCG	85
QY	112	cttcattacatacagatgcgcgcttttcgcacacagtcagaatatgtagaccgcttcg	171

Db	86	 TTTCACTACATACAGATGGCGCTTTTGGCACACAGTCAGAATATGCAGATGCCTTGC	145
Qy	172	tgctgcaattgaagcagaaaaatcaattgacaagtgcagccagagattccttattgacaagat	231
Db	146	TGCTGCAATTTGAAGCAGAGAAATCACTGTACAAGTCAGACCAGATTGCTATTGCACCAT	205
Qy	232	tagccagcagcaaggaagtagtagctcttgaagaacaaatgaaagcatcagagaccagga	291
Db	206	TAGCCAGCAGCAAGGAAGAATAGTTGCTCTTTGAAGAAACAAATCAAGCGTCAGGACCAGGA	265
Qy	292	gtgcgcgcaattagggctctgttcaggatcttgaagtaagggcataaaaaagttlaa	351
Db	266	GTGCCGACAGTTAAGGCTCTGTGTCAGGATCTTTGAAGGTAAAGGCATAAAAAAGTTGAT	325
Qy	352	cggagatgtgcagatgcgaagtggcagctgtagttgttatgtgttcgctgcagtcgtactgacta	411
Db	336	CGGAAATGTCACAGATGCCAGTGCGTGTAGTTGTTATGGCTTGCAATCGGCTGACTA	385
Qy	412	cctggagaggactataaatccatctttaaaatcccaaaacatctgttgcatacaaaatccc	471
Db	386	CCTGGAAAGACTATTAAATCCATCTTTAAATACCAAAATATCTGTTCGCCAAAAATATCC	445
Qy	472	tctttccatatcccagatggatcaaatcctgatglaagaaagcttgctttgagctatgg	531
Db	446	TCCTTTTCATATCCACAGATGGATCACAATCCTGTATGTTAGCAAGCTTGCTTTGAGCTATGA	505
Qy	532	tcagctgacgtatatgcagcaacttggattatgaacctgtgtcatactgaaagaccagggga	591
Db	506	TCAGCTGACGTATATGCAGACACTTGGATTTTGAACCTGTGCATCTGAAAGACCAGGGGA	565
Qy	592	actggttgcaactacaagattccaagtcattacaagtgggcattgagtcagctgtttca	651
Db	566	GCTGATTGCATCTACAAAATTGACGCTCAATTACAAGTGGCATTTGGATTCAGCTGTTTGA	625
Qy	652	caagcataaatttagccgtgtctatcatactagaagatgatatggaaattgctgctgattt	711
Db	626	CAAGCATAAATTTAGCCGTGTTATCATACTAGAAGATGATATGGAAATTTGCCCTGATTT	685
Qy	712	ttttgactatttgagctggaagctactcttcttgacagagacaagtcgattatagctat	771
Db	686	TTTTGACTTTTGTGAGCTGGAGCTACTCTTCTGTGACAGAGACAAGTCGATTATGGCTAT	745
Qy	772	ttctcttggaaatgacaatggacaaaggcagttcgttccaagatccctgaigtctcttaccg	831
Db	746	TTCTTCTTGAATGCACAATGCACAAATGCAGTTTGTTCCAAGATCCTTATGTCTCTTACCG	805
Qy	832	ctcagaactttttcttgcttgctcttgatggatgctttcaaaatcaactggccgaactatc	891
Db	806	CTCTGTATTTTTTCCCGGCTCTGGATGGATGCTTTCCAAAATCTACTTGGGACGAACATC	865
Qy	892	tccaaagtgcgcaaaagccttactgggatgactgctgaagcgtgaagaaaaatccagaggg	951
Db	866	TCCAAAGTGCCAAAGGCTTACTTGGGACGACTGGCTTAAGACTCAAGAGAAATCAGAGG	925
Qy	952	tcgacaatttatcgccagaagcttgcagaaacgtacaaatttggctgagcatggtcttag	1011
Db	926	TCGACAAATTTATTGGCCAGAAAGTTTGCAGATCATATAATTTTGTGTGAGCATGTTCTTAG	985
Qy	1012	tttggggcagtttttaagcagtatcttgagccaattaagctaaaatgatgtccaggttga	1071
Db	986	TTTGGGGCAGTTTTTCAAGCAGTATCTTGAAGCCAATTAACCTAAATGATGTCCAGGTGA	1045
Qy	1072	ttggaagtcaatggacctaagttacccttttgaggacaactatgtgaacacttttgoga	1131
Db	1046	TTGGAAGTCAATGGAGCTTAGTTTACCCTTTTGGAGGACAAATACGTTGAACACTTTTGGTGA	1105
Qy	1132	cttggttaaaagcctaagcccatccagagagctatgctgttttgaagcatttaacat	1191
Db	1106	CTTGGTTAAAGAGCTTAAGCCCATCCATGGAGCTGATGCTGTTTGTGAAGCATTTTACAT	1165
Qy	1192	agatggtgatgtcgctatttcagtacagagacccaactagacttttgaagatatctgcctgaca	1251

Db	1166	AGATGGTGATGTGCGTATTTCAGTAGACAGAGATCAACTAGACTTTGAAGATATCGCACGGCA	1225
QY	1252	qtttggcatttttgaagaalggagatggtgtaccacgagcagcatataaaggatagt	1311
Db	1226	ATTTCGATTTTTCAGAAATGCGAGGATGGTGTACCGAGGACATATAAGGAATAGT	1285
QY	1312	agttttccggtttcaaacatctagacagtggtgtcttctgtttcccttgattctctcgaca	1371
Db	1286	GGTTTTCGGGTACCAACAGCCAGACGTGATTCTCTGTTGGCCCTGATTCGCTTCAACA	1345
QY	1372	acttggagttgaagatacttagcgaagaatagattgagccttagcgaacaatttagactt	1431
Db	1346	ACTCGGAATTAAGAGATACTTAAACAAGATATGATTGGAGCCCGACAAAGATTTAGACTT	1405
QY	1432	atttggtaggatacatttgaagagctgacacgaagagtagtactacagtagctacatg	1491
Db	1406	ATTGGGTAGGCCA---TGATCGAGCTGACACCAACCATGAGTT--TACAGTTACATA	1460
QY	1492	caacattftaa-tgttaattggaaggaaacccactgcttattgttggatgat	1542
Db	1461	CAACGTTTAAATGTTATATGGAGGAGCTCACTGTTCTAGTGTGAAGGAT	1512
RESULT 9			
ATH243198			
LOCUS			
DEFINITION			
Arabidopsis thaliana mRNA for N-acetylglucosaminyltransferase I			
(GlcNAc-T I gene).			
ACCESSION			
AJ243198			
VERSION			
AJ243198.1 GI:5139334			
KEYWORDS			
GlcNAc-T I gene; glycosyl transferase;			
N-acetylglucosaminyltransferase I.			
SOURCE			
thale cress.			
ORGANISM			
Arabidopsis thaliana			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
1 (bases 1 to 1820)			
Bakker,H., Lommen,A., Jordi,W., Stiekema,W. and Bosch,D.			
An arabidopsis thaliana cDNA complements the			
N-acetylglucosaminyltransferase I deficiency of CHO lec1 cells			
Biochem Biophys. Res. Commun. 261 (3), 829-832 (1999)			
99373163			
2 (bases 1 to 1820)			
Bakker,H.			
Direct Submission			
Submitted (17-JUN-1999) Bakker H., Molecular Biology, Centre for			
Plant Breeding and Reproduction Research, P.O. Box 16 Wageningen,			
6700AA, NETHERLANDS			
Related sequence AL035538.			
COMMENT			
FEATURES			
source			
1. .1820			
/organism="Arabidopsis thaliana"			
/variety="Colombia"			
/db_xref="taxon:3702"			
/clone="pABE-ALGN1"			
/clone_lib="pABE-A.t.seliq"			
31. .1365			
/gene="GlcNAc-T"			
31. .1365			
/gene="GlcNAc-T"			
/EC_number="2.4.1.101"			
/function="glycosyltransferase"			
/codon_start=1			
/product="N-acetylglucosaminyltransferase 1"			
/protein_id="CAB45521.1"			
/db_xref="GI:5139335"			
/translation="MARISDLRFLIPAEWFYIQMLFQTSQVADRUSIAIESE			
NHCISQMKGLIDEYSIKQSRIVALDEMKNRQDELVLQKDLIQTFKKGIKAKLQGGQ			
MPVAAYVMVMSRADYLERTVKSVLTQTVPVASKYPLFISODGSDQAVKSKLSYNOL			
TYMQHLDPEFVYTPERPGLTAYKIARHYKWLQDLFYKHKFSRVIILEDDEMEIAPDF			
FDYFEAAASLMDRDKTTMAASSNDNGKQFVHPYALYSDFPFPGWMLKRWIDE			
LSPKPKAYWDDWLRLKENHKGQRFIRPEVCRTYNFGEHSSSLGQFFSQYLEPIKIND			

BASE COUNT		532 a	342 c	421 g	525 t
ORIGIN					
Query Match		45.1%	Score 752.6;	DB 8;	Length 1820;
Best Local Similarity		72.7%;	Pred. No. 5.2e-179;		
Matches 971;		Conservative 0;	Mismatches 364;	Indels 0;	Gaps 0;
QY	65	aagtttgcttgattacggtgaccttcctgctgctgctgctcctcgccttcacata	124		
Db	37	AGGATCTCGTGACTTGAGATTCTTCTCATCCGCGCAGCTTTCATGTTTCATCTACATC	96		
QY	125	cagatcgcttttcgcgacacagtcagaataatgtagaccgcttgcgtgcgaattgaa	184		
Db	97	CAGATGAGGCTTTTCCAGACGCAATCACAGTATGCAGATGCCTCAGTTCGGTATCGAA	156		
QY	185	gcagaaaaatctgtacaagtcagaccagattgcttattgacaagaattagccagacaa	244		
Db	157	TCTGAGAACCATTTGCACATAGTCAAAATCCGAGGCTCATAGATGAAGTTAGCATCAACAG	216		
QY	245	ggaagagtgtgctcttgagaacaaaatgaagcatcagaccagagtgccggcaatta	304		
Db	217	TCGCGGATTGTTGCCCTCGAAGATATGAAGAACCGCCAGGACGAAGAACTTGTGCAGCTT	276		
QY	305	agggctctgttcagagatcttgaagaatgaaggcataaaaaagttaatcggagatgtgcag	364		
Db	277	AAGGATCTAATCCAGACAGCTTTGAAAAAAGAAATAGCAAACTCAGTCAAGGTGGACAG	336		
QY	365	atgccagtgccagctgtagttgtatggcttcagtcgctactacacctggagagagact	424		
Db	337	ATGCCCTGGCTGCTAGTGGTTATGGCCCTGCAGTCGTGCAGACTATCTTGAAGGACT	396		
QY	425	attaaatccatcttaaaataaccaaacatctgttgcatcaaaaatctcctctttcatatcc	484		
Db	397	GTTAAATCAGTTTTTAACATATCAAACTCCCGTTGCTTCAAAATATCCTCTATTATATCT	456		
QY	485	caggatgatcaaatcctcgtatgaagaagcttgccttgagctatggctcagctgacgatat	544		
Db	457	CAGGATGGATCTGATCAAGCTGTCAAGAGCAAGTCATTTGAGCTATAATCAATTAACATAT	516		
QY	545	atgcagcacttgattatgaacctgtgcatactgaagaccaggggaactgggtgcatac	604		
Db	517	ATGCAGCACTTGGATTTTGACCCAGTGGTCACGTGAAGGCTCGCGAACTGACTGGCTAC	576		
QY	605	tacaagatgcagctcattacaagtgggcatttgatgcagctgtttcacacagcataatttt	664		
Db	577	TACAAGATTGCAGTCACTACAAAGTGGGCACATGGACCAGTTGTTTACAAACACAAATTT	636		
QY	665	agccgtgttatcatatacagaagatgatagaaattgctgctgattttttgactatttt	724		
Db	637	AGTCGAGTGATTATACTAGAAGATGATATGGAAATTTGCTCCAGACTTCTTTTATTAATCTT	696		
QY	725	gaggtcgagctactcttctcacagacacagtcgattatggctattctctctcttggaa	784		
Db	697	GAGGCTGCAGCTAGTCTCATGGATAGGGATAAAACCATTTATGGCTGCTTCATCATGGAAT	756		
QY	785	gacaatggacaaggcagtttcgaagatcctgctgctgtcttttacgcgtcagactttttt	844		
Db	757	GATAATGGACAAAGCAGTTTGTGATGATCCTATGGCTATACCGATCAGATTTTTTT	816		
QY	845	cctggtcttgatggatgctttcaaaatacaacttggccgaactactctccaaagtggcca	904		
Db	817	CCTGGCCTTGGGTGGATGCTCAAGAGATCGACTTGGGATGAGTTATCAACAAAGTGCCA	876		
QY	905	aaggttactgggatgactggctgagctgaaggctgaagaaaaacacagaggtcgacaaatttt	964		
Db	877	AAGGCTTACTGGGATGATTGGCTGAGACTTAAGGAAAAACCATTAAGAGCGCGCAATTCATT	936		
QY	965	gcgccagaagtttgcagaacgtacaaattttggtgagcagttgtctagtttggggcagttt	1024		
Db	937	CGACCGGAAGTCTGTAGAACAATACAATTTTGGTGAACATGGGTCTAGTTTGGGACAGTTT	996		



Db	963	AAGGCTTACTGGGATGATTGGCTGAGACTAAAGGAAAAACCAATAAAGGCCGCCCAATTCATT	1022
Qy	965	cgccagaaagttagcagaacgtacaaatttggtagcatggttcttaagttagtgaggcaagttt	1024
Db	1023	GCACCGGAAGTCTGTAGAACAATAACAATTTTGGTCAACATGGGTCTAGTTTGGGACAGCTTT	1082
Qy	1025	ttaaagcagtattcttagccaataaagcctaagcgaatgagtcagagtgatgattggaagcgaatg	1084
Db	1083	TTCACTCAGTATCTGGAACTATATAAGCTAAACGATGTGACGGTTGACTGGAAGCAAG	1142
Qy	1085	gacctaaagttacatttttggaggacaactatgtgaacacatttggcgacttggttataaaag	1144
Db	1143	GACCTGGGATACCTGCACAGAGGAAACATATACCAAGTACTTTTCTTGCTTAGTGAGACAA	1202
Qy	1145	gctaagcccataccagagcgtgactgctgttttgaagcattttaacatagatgtagtg	1204
Db	1203	GCAGCACCAATTCAAGGTTCTTGACTTGTCTTAAGGCTCAAAACATAAAGATGATGAT	1262
Qy	1205	cgtattcagtcacagagaccactagactttgaagatatacgtcgcagacagtttggcattttt	1264
Db	1263	CGTATCCGGTATAAAGACCAAGTAGAGTTTGAACGCATTGCAGGGGAATTTGGTATATTT	1322
Qy	1265	gaagaatggaagtagtggttacacagggcagcgcataataaaggagatagtagtttccggttt	1324
Db	1323	GAAGAATGGAAGGATGCTGTGCCACGACGACATATAAAGGAGTAGTGGTGTTCGGAATC	1382
Qy	1325	caaacactagacgtgtgttcctgttccctgattctcttcgcacacacttgaggttga	1384
Db	1383	CAGACACACAGACGTGTATTCCTGGTGGGCCGACATTCTGTAAATGCAGCTTGGNAATCGA	1442
Qy	1385	gatacttagcgaaga	1399
Db	1443	AATTCCTGATGCAAA	1457
RESULT	11		
LOCUS	A95057		
DEFINITION	Sequence 5 from Patent WO929879.	1854 bp	DNA
ACCESSION	A95057		linear
VERSION	A95057.1	GI:6779211	PAT 26-JAN-2000
KEYWORDS			
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 1854)		
AUTHORS	Von, S. A.		
TITLE	VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH		
JOURNAL	A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GntI)		
	ACTIVITY		
	Patent: WO 9929879-A 5 17-JUN-1999;		
FEATURES	VON SCHAEWEN ANTJE (DE)		
source	Location/Qualifiers		
	1..1854		
	/organism="Arabidopsis thaliana"		
	/strain="COLOMBIA"		
	/db_xref="taxon:3702"		
	/tissue_type="ALLE GEWEBE"		
	/clone_lib="LAMBDA UNI-ZAP (ECORI/XHOI) UND LAMBDA ACT"		
	/clone_lib="(XHOI)"		
	/clone_lib="(B) CLON(E): PBSK-ARA-GNTI-FULL #8"		
	/dev_stage="M: REIFE PFLANZEN"		
	19..134		
	135..1469		
5'UTR	/gene="CGL"		
gene	135..1469		
CDS	/gene="CGL"		
	/standard_name="GNTI"		
	/EC_number="2.4.1.101"		
	/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.		
	GLYKOPROTEINEN"		

/note="ERSTE GNTI-SEQUENZ AUS ARABIDOPSIS (UNPUBLIZIERT)"			
/codon_start=1			
/evidence=experimental			
/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I"			
/protein_id="CAB70466.1"			
/db_xref="GI:6779212"			
/translation="MARISCDLRFLLIIPAAFMFIYIQMLRFQTSQYADRLSSAIESE			
NPTVSOMRGLIDEVSIKQSRIVALEDMKNRQDEELVOLKDIQTFFKKGIAKLITQGGQ			
MHVAAVVMACSRADYLERIVASLVLTQYTPVASKYPLFIQSDGSDQAVKSKLSYNOL			
TYQHLDLFEFVYTERGELTAYKIKARHYKWDOLFYKHKFSRVLILEDDETAPDF			
FDFAAASLMDRDKTITMASSNNDNGQKQFVHDHPVALIRSDRFFGLGLMKRSTWDE			
LSPKWPAYMDMLRLKENHKQFIAPFCRPIQFSGSLGQFSSLGQFFSQWLEPKLMD			
VIVDMKAKDLGYLTGNTYKFSGLVRQARPIQGSDLVLKAQNIKDDRIRYKDQVEF			
ERLAGEFGIFEKWKDGVPRYAYKGVVVVFRIQTRRVFLVGPDSVMQLGIRNS"			
<157..>215			
/gene="CGL"			
/standard_name="MEMBRANANKER EINES TYP II GOLGI-PROTEINS"			
/function="MEMRANKER (AMINO ACID 8-27)"			
/note="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN			
IDENTIFIZIERT"			
/codon_start=2			
/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"			
/protein_id="CAB70467.1"			
/db_xref="GI:6779213"			
/translation="EISSHPGSPHVLHPDEAF"			
1470..1848			
3'UTR	545 a	357 c	430 g 522 t
BASE COUNT			
ORIGIN			
Query Match 44.8%; Score 747.8; DB 6; Length 1854;			
Best Local Similarity 72.5%; Pred. No. 8.4e-178;			
Matches 968; Conservative 0; Mismatches 367; Indels 0; Gaps 0;			
Qy	65	aagtttgcttaattacggtacattctcgctgctgctctcgccttcacatcata	124
Db	141	AGGATCTCGTGTACTTGAGATTTCTTCTCATCCCGCAGCTTTCATGTCATCATC	200
Qy	125	cagatcgcgcttttcgcgcacagtcagaataatgtacccgcttctgctgcaattgaa	184
Db	201	CAGATGAGGCTTTTCCAGACGCAATCACAGTATGCAGATCGCCTCAGTTCGCTATCGAA	260
Qy	185	gcgagaaatcattgtacaagtcagaccagattgcttattgacaagattagccagacaa	244
Db	261	TCTGAGAACCATTCGACTAGTCAAAATCGAGCGCTCATAGATGAAGTTAGATCAAAACAG	320
Qy	245	ggaagagtagtagctcttgaagaacaaatgaagcatcagaccagaggtgcccgaatta	304
Db	321	TCGCGGATTGTTGCCCTCGAAGATATGAAGAACCGCCAGGACGAAGAACTTGTCCAGCTT	380
Qy	305	agggctcttctcaggatcttgaaagttaaggcataaaaaagttaatcggagatgtgcag	364
Db	381	AAGGATCTAATCCAGACGCTTTGAAAAAAGGAATAACAAAACCTCACTCAAGGTGGACAG	440
Qy	365	atccagtcgacgttagttgttattgcttcagctcgactgactactacclggagagact	424
Db	441	ATGCTGTGGCTGCTGTAGTGGTTATGGCCTGCGACTCGTCGAGACTATCTTGAAGGACT	500
Qy	425	attaaatccatcttaaaataccaaacatctgttgcatcaaaaatacctcttttcatatcc	484
Db	501	GTTAAATCAGTTTTAACATATCAAACTCCCGTTCCTTCAAAATATCTCTATTTATATCT	560
Qy	485	caggatggatcaaatcctgtagtaagaagcttgcctttagctatggtcagctgacgtat	544
Db	561	CAGGATGGATCTGATCAAGCTGTCAAGAGCAAGTCATTGAGCTATAATCAATTAACATAT	620
Qy	545	atgcagcaacttgattatgaacctgtgtcatcaciagaagaccaggggaactggttgctaac	604
Db	621	ATGCAGCACTTGGATTTTGAACAGTGGTCACTCAAAAGGCTTGGTGAACCTGACTGCGTAC	680
Qy	605	tacaagattgcagcgtcattacaagtgggcattggatcagctgttttcacagcgtataattt	664
Db	681	TACAAGATTGCAGTCACTACAAGTGGGCACCTGGACCAGTGTGTTTTACAAACACAAATTT	740

QY	665	agcgtgttatcatatactagaagaatgatatggaattgctgctgatttttttgactatttt	724
Db	741	AGTCGAGTGATTATACTAGAAGACCATATGGAATTTGCTCCAGACTTCTTTGATTACTTT	800
QY	725	gaggctggaactactcttcttgacagagacaagaatcgattatgctattcttctcttggaaat	784
Db	801	GAGCTGCAGCTAGTCTCATGATAGGATAAAACCATTTATGGCTCTCTCATCATGGAAT	860
QY	785	gacaatggacaaaggcagttcgtccaaagatccctgatgctctctttaccgctcagactttttt	844
Db	861	GATAATGGACAGAAGCAGTTTGTGCATGATCCCTATGCGCTATACCGATCAGATTTTTT	920
QY	845	cttggtcttgatgtagtcttccaataccaacttgctgcgaactatctccaagttggcca	904
Db	921	CCTGGCCTTGGTGGATGCTCAAGAGATCGACTTGGATGAGTTATACCAAAATGGCCA	980
QY	905	aagcttactggatgactgtgctaaagctgaaagaataacacagagtgctacaattatt	964
Db	981	AAGCCTTACTGGGATGATTGGCTGAGACTAAAGAAACCATAAAGCCGCCAATTCATT	1040
QY	965	cgccagaagtcttgacagaacgtacaatttttggtagcatggttctagtttggggcagttt	1024
Db	1041	GCACCGGAAGTCTGTAGAACATACAATTTTGGTGAACATGGTCTAGTTTGGACAGTTT	1100
QY	1025	titaagcagtatctdgaccaattaagctaaatgatgtccaggttgattggaagtcaatg	1084
Db	1101	TTCACTCAGTATCTGGAACCTATTAAGCTAAACGATGTGACGGTTGACTGGAAGCAAG	1160
QY	1085	gacctaaagtacccttttggaggacaactatgtgaaacaacttttggcacttgggttaaaag	1144
Db	1161	CACCTGGATACCTGCAGAGAGGAAACTATACCAAGTACTTTTCTGGCTTAGTCAGACAA	1220
QY	1145	gctaagcccatccagagcgtgctgttttgaagcatttaacatagatgctgatgtg	1204
Db	1221	GCAGACCAATTCAGGTTCTGACCTTGTCTTAAAGCTCAAAACATPAAAGGATGATGAT	1280
QY	1205	cgtattcagtagacagaccactagactttgaaagatatcgtcgacagtttggcattttt	1264
Db	1281	CGTATCCGGTATAAAGACCAAGTAGAGTTTGAACGGCATTCAGGGGAATTTGGTATATTT	1340
QY	1265	gaacaatggaagtagtgttaccagggcagcatataaaggatagtagtttcccggtt	1324
Db	1341	GAAGAAATGGAAGTAGTGTGGCCAGCAACGACATATAAAGAGTAGTGGTTCGGAATC	1400
QY	1325	caaacatctagacgtgtgttccttctgttccctgattctctctcgacaacttggagttgaa	1384
Db	1401	CAGACAACAGACGTGTATTCTCTGGTTGGCCACAGATTCTGTATGTCAGCTTGGAAATCGA	1460
QY	1385	gatacttagcgaaga	1399
Db	1461	AATTCCTGATGCAAA	1475
RESULT 12			
MUSTRANSX			
LOCUS	MUS musculus (clone E31.1 in pGEM7zf(+))		
DEFINITION	N-acetylglucosaminyltransferase I mRNA, complete cds.		
VERSION	L07037		
KEYWORDS	L07037.1 GI:202145		
SOURCE	N-acetylglucosaminyltransferase I.		
ORGANISM	Mus musculus		
	of Robert Larsen (strain BALB/c, sub-species domesticus) (library: cDNA		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2695)		
	Kumar,R., Yang,J., Eddy,R.L., Jr., Byers,M.G., Shows,T.B., Jr. and Stanley,P.M.		
TITLE	Cloning and expression of the murine gene and chromosomal location		
	of the human gene encoding N-acetylglucosaminyltransferase I		
JOURNAL	Glycobiology 2, 383-393 (1992)		

MEDLINE	93043841	Location/Qualifiers
FEATURES		
source		
		1. .2695
		/organism="Mus musculus"
		/strain="BALB/c"
		/sub_species="domesticus"
		/db_xref="taxon:10090"
		/cell_line="F9 embryonal carcinoma"
		/tissue_lib="stem cell"
		/tissue_lib="cDNA of Robert Larsen"
CDS		364. .1707
		/codon_start=1
		/product="N-acetylglucosaminyltransferase I"
		/protein_id="AAA40478.1"
		/db_xref="GI:202146"
		/translation="MLKQTAGLVLMGALIFVGNALLLLFFWTRPAPGRRLSDSALG
		DDPSLSTREVIHLAEADAELERQLQOIKEHYALMQRWRVPIVAPMPRVPT
		VTHIQPDLISIAVDPHRKFOGYKIHRYWALQOIFNKFFPAAVVEDLEVP
		DFFEVFOATYPLLRTPDPSLMCVSANDNGKEQVDSKPELLYRTFFPLGLMLLAD
		LWNELEPKWPKAFWDDMMERPEQRGRACIRPEISRTMTFGRKGVSHGFQDHLKFI
		KLNQDFVPTQLDLSYLQGEAYDRFLAQYVGAPOLOVEKVRTNDOKELGEVRVOYTS
		RDSFNAFAKALGVMDLKSQVPRAGTIGIVTFQRRRVHLAPPQTWTGYDPSWN"
terminator		1705
3'UTR		1708. .2695
polyA_site		2695
BASE COUNT	546 a	715 c 780 g 654 t
ORIGIN		

Query Match	12.1%	Score	202.4;	DB	10;	Length	2695;
Best Local Similarity	55.9%	Pred. No.	2.6e-40;				
Matches	408;	Conservative	0;	Mismatches	316;	Indels	6;
Gaps	1;						
QY	602	tactacaagattcacagctcattacaagtgggcattgagcagctgtttccacaagcataat	661				
Db	913	TACTACAAGATTGCCAGGCACCTACCGCTGGCCTAGGCCAGATCTTCAACAAGTTCAAG	972				
QY	662	tttagccgtgttatcactactagaagatgatatggaattgctgtgatttttggactat	721				
Db	973	TTCCCGCCGCTGTGGTAGTGAGGATGATCTGGAAGTGGCACCAGACTTCTTTGAGTAC	1032				
QY	722	tttgagcgtgagcactactctttgacagagacaagtcgattatgctatttcttcttgg	781				
Db	1033	TTCCAGGCCACCTACCCACTGCTGAGAACACAGACCCCTTGGTGTGTGCTGTGGTTGG	1092				
QY	782	aatgacaatggacaaaggcagttcgtcca-----agatcctgatgctctttaccgctca	835				
Db	1093	AATGACAAATGGCAAGGACAGATGGTAGACTCAAGCAAACTGAGCTGCTTATCGAACA	1152				
QY	836	gacttttttccctggttggatgtagcttttcaaaatcaaaacttggtccgaactatctcca	895				
Db	1153	GACTTTTCTGCGCTTGGATGCTGTGTGGCTGATCTGTGGCGACAGCTGGAGCCC	1212				
QY	896	aagtggccaaaggcttactgggatgactggcctaaggctgaaagaaatcacagaggtcga	955				
Db	1213	AAAGTGGCCCAAGGCTTTTGGGACGACTGGATGCGCCGACCTGAGCAGCGAAAGGACGG	1272				
QY	956	caatttatttcgccccagaagtttgcagacgtacaaattttggtgagcagtggttctagtttg	1015				
Db	1273	GCATTGATTCTGCCAGAAATTTCAAGAACTATGACCTTTGGTCGCAAGGCTGTGAGCCAC	1332				
QY	1016	gggcagtttttaagcagtagtatcttgaccgaataaagcctaagcaggttccagggttattg	1075				
Db	1333	GGGCAGTCTTTGACCAAGCATCTTAAGTTCATCAAGCTGAACACGACGAGTTCGTCCTTC	1392				
QY	1076	aagtcaatggacctaaagttaaccttttggaggacaactatgtgaaacaacttggcgacttg	1135				
Db	1393	ACCCAGTTGACACCTGCTGCTACCTGCACGACGAGGCCCTATGACCGGACTTCTTCGCCAC	1452				
QY	1136	gttaaaaggcctaagcccatccacggagctgtagtctgttttgaagcatttaacatagat	1195				
Db	1453	GTCTATGGTGGCCCCCAGCTACAGGTGGAGAAAGTAAGGACCAATGATCATCAAGAGGAGCTG	1512				







```
misc_feature 1. 6214
/note="assembly_fragment:05996
fragment_chain:1"
misc_feature 6315..10914
/note="assembly_fragment:05775
fragment_chain:1"
misc_feature 11015..14003
/note="assembly_fragment:06421
fragment_chain:2"
misc_feature 14104..16195
/note="assembly_fragment:06537
fragment_chain:2"
misc_feature 16296..219581
/note="assembly_fragment:04325"
misc_feature 219682..233789
/note="assembly_fragment:05835"
misc_feature 233890..240341
/note="assembly_fragment:05852"
misc_feature 240442..244472
/note="assembly_fragment:05915"
misc_feature 244573..261705
/note="assembly_fragment:06317"
misc_feature 261806..267628
/note="assembly_fragment:06492"

BASE COUNT 75659 a 55085 c 56023 g 79959 t 902 others
ORIGIN
```

```
Query Match 12.0%; Score 200.8; DB 2; Length 267628;
Best Local Similarity 55.8%; Pred. No. 6.2e-40;
Matches 407; Conservative 0; Mismatches 317; Indels 6; Gaps 1;

QY 602 tactacaagattgcagctcattacaagtgggcattggatcagctgtttcacagaataat 661
|||||
Db 126195 TACTACAAAGATTGCCAGGCACCTACCGCTGGGCACCTAGGCACAGATCTCAACAAGTTCAAAG 126254
|||||

QY 662 tttagccgtgttatcatactagaagaatgatggaattgctgctgattttttgaactat 721
|||||
Db 126255 TTCCCGCGCGTGTGGTAGTGAGGATGATCTGGAAGTGGGCACACGACTCTTTGAGTAC 126314
|||||

QY 722 tttagagctggagctactcttcttgacagagacaagtcgattatggctattctctcttgg 781
|||||
Db 126315 TTCCAGGCCACCTACCCACTGCTGAGAACAGACCCCTCCCTTTGGTGTGTGCTCTGCTTGG 126374
|||||

QY 782 aatgacaatggacaagcaggtctcca-----agatcctgatgctcttttaacgcctca 835
|||||
Db 126375 AATGACAATGGCAAGGACGAGATGTAGACTCAAGCAAAACCTGAGCTGCTCTATCGAACA 126434
|||||

QY 836 gactttttcctggtcttggatggatgctttcaaaatcaaaacttggtcgaaactatctcca 895
|||||
Db 126435 GACTTTTTCTGTGGCCTTGGATGGCTGCTGTGTGGTGATCTGTGGCAGAGCTGGAGCCC 126494
|||||

QY 896 aagtggcaaaaggcttaactgggatgactggctaaagctgaaagaaaaatcacagaggtcga 955
|||||
Db 126495 AAGTGGCCCAAGGCCCTTTTGGGACGACTGGATGCCCGCACCTGAGCAGCGGAAGGAGCGG 126554
|||||

QY 956 caatttattcgccagaagtttgacagacgtacaaattttggtgagcatggttcttagtttg 1015
|||||
Db 126555 GCTTGATTCGTCCAGAAATTTCAAGAACTATGACCTTTGGTCGCAAGGGTGTGAGCCAC 126614
|||||

QY 1016 gggcagtttttaagcagatatttgagccaaattaaagctaaatgatgtccaggttgattgg 1075
|||||
Db 126615 GGGCAGTTCTTTTGACCAAGCATCTTAAGTTCATCAAGCTGAACGACAGTTTCGCCCTTC 126674
|||||

QY 1076 aagtcattggaccttaagctttttgagacaaactatgtgaaacacttttggcacttg 1135
|||||
Db 126675 ACCCAGTTGGACCTGCTGCTACCTTGAGGAGGAGGCTATGACCGGGAGCTTCCTCGCCCGAG 126734
|||||

QY 1136 gtaaaaaagcctaagccctccacggagctgatgctgtttttgaaagcatttaacatagat 1195
|||||
Db 126735 GTCTATGTTGCCCCCCAGCTACAGGTGGAGAAAGTAAGGACCAATGATCAGAAAGGAGCTG 126794
|||||

QY 1196 ggtgatgtgcgtatttcagtagacagaccactagactttgaagatatcgtctcgacagttt 1255
|||||
```

```
Db 126795 GGGGAGGTCCGGGTACAGTACACTACAGACAGACAGCTTCAAGGCCCTTGTGTAAGGCCCTG 126854
|||||
QY 1256 ggcatttttgaagaatggaagatggtgtaccacgggcagcatatataaaggatagtagtt 1315
|||||
Db 126855 GGTGTTCATGGATGACCTCAAGTCTGGTGTCCCGAGGGCTGGCTACCGTGGCATTTGCACT 126914
|||||
QY 1316 ttccggtttc 1325
|||||
Db 126915 TTCCAGTTTC 126924
|||||
```

Search completed: August 13, 2002, 20:28:36  
Job time: 9:59 sec



OM of: US-09-591-466c-2 to: EST:\* out\_format : pfs  
Date: Aug 13, 2002 10:41 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame\*\_p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09591466/runat\_12082002\_095312\_4249/app\_query.fasta\_1.511  
-DB=EST -Qfmt=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPOP=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09591466 -CGN1\_1.4902  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-591-466c-2  
Query length: 446  
Database: EST:\*  
Database sequences: 13736207  
Database length: 1841457050  
Search time (sec): 1757.600000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est2:BG594788	+	1143.00	2310.95	1.4e-120	768
gb_est1:AI486771	+	1116.00	2277.34	1.3e-117	643
gb_est2:BG592816	+	1113.00	2271.18	2.9e-117	642
gb_est2:BI1779808	+	1112.00	2267.75	4.6e-117	724
gb_est1:AI725875	+	1020.00	2078.90	1.5e-106	681
gb_est2:BG889872	+	1012.00	2062.76	1.2e-105	661
gb_est1:AW931180	+	988.00	2014.91	5.5e-103	574
gb_est1:AW979500	+	974.00	1984.26	2.8e-101	673
gb_est1:AW735827	+	954.00	1943.29	5.4e-99	659
gb_est2:BE426174	+	953.00	1941.76	6.5e-99	629
gb_est2:BE516817	+	859.00	1748.57	3.8e-88	503
gb_est2:BG409413	+	832.00	1693.17	4.6e-85	591
gb_est2:BE805530	+	825.00	1677.43	3.5e-84	664
gb_est2:BG726055	+	764.50	1555.86	2.0e-77	506
gb_hic:AK004760	+	732.00	1469.97	1.2e-72	2681
gb_est2:BJ170544	+	724.00	1468.72	1.5e-72	701
gb_est2:BG508571	+	683.00	1388.36	4.4e-68	488
gb_est2:BF066219	+	669.50	1362.30	1.2e-66	418
gb_est2:BM177691	+	662.00	1344.26	1.3e-65	525
gb_est2:BF145860	+	642.00	1300.90	3.3e-63	635
gb_est2:BG680531	+	631.00	1274.35	9.8e-62	894
gb_hic:AF289558	+	608.50	1216.04	2.7e-58	2561
gb_est1:AW618699	+	608.00	1232.90	2.0e-59	530
gb_est2:BG051233	+	601.00	1219.79	1.1e-58	472
gb_est2:C96708	+	586.00	1183.54	1.1e-56	758
gb_est2:BG744010	+	572.50	1155.11	4.3e-55	798
gb_est2:BM137146	+	561.50	1136.31	4.8e-54	568
gb_est2:BE404993	+	561.00	1134.89	5.7e-54	588
gb_est2:BG131020	+	561.00	1134.62	5.9e-54	602
gb_est2:BI152075	+	557.50	1124.16	2.3e-53	801
gb_est2:BG682233	+	548.50	1104.35	2.9e-52	896
gb_est1:AW670629	+	531.00	1090.40	2.7e-51	546
gb_est2:BG917557	+	539.00	1089.26	1.6e-50	823
gb_est1:AL503600	+	529.50	1068.17	3.0e-50	690
gb_est2:BE412855	+	529.50	1068.17	3.0e-50	690
gb_est2:BI415357	+	529.00	1065.29	4.3e-50	889
gb_est2:BG680449	+	528.00	1064.22	5.0e-50	744
gb_est2:BE289472	+	517.00	1042.63	7.9e-49	677
gb_est2:BG481478	+	515.00	1035.81	1.9e-48	859
gb_est1:AI897249	+	507.00	1024.85	7.7e-48	528

gb\_est2:BM427482 + 506.00 1021.02 1.3e-47 617 i BM427482 pgf2n.pk006.ol8 Nor  
gb\_est2:BE821775 + 496.00 999.13 2.1e-46 691 i BE821775 GM700015A20F4 Gm-r1  
gb\_est1:AU179654 + 495.00 1000.75 1.7e-46 500 i AU179654 AU179654 Medaka liv  
gb\_est2:BG888733 + 484.00 978.36 3.0e-45 488 i BG888733 EST514584 cSTD Sola  
gb\_est2:BE382846 + 479.00 963.72 2.0e-44 715 i BE382846 601297718F1 NIH\_MGC  
seq\_name: gb\_est2:BG594788  
seq\_documentation\_block:  
LOCUS BG594788 768 bp mRNA linear EST 12-APR-2001  
DEFINITION EST493466 cSTS Solanum tuberosum cDNA clone cSTS8020 5' sequence,  
mRNA sequence.  
ACCESSION BG594788  
VERSION BG594788.1 GI:13612928  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Solanum.  
REFERENCE 1 (bases 1 to 768)  
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,  
Bougril,O., Buell,C.R., Romning,C., Tanksley,S. and Baker,B.  
Generations of ESTs from sprouting potato eyes  
UNPUBLISHED (2000)  
CONTACT: Cathy Ronning  
The Institute for Genomic Research  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email cdna@resgen.com  
COMMENT  
Seq primer: M13F-R.  
FEATURES  
Location/Qualifiers  
1..768  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone="cSTS8020"  
/clone.lib="cSTS"  
/dev\_stage="12-14 weeks post harvest"  
/lab\_host="SOLR"  
/note="vector: pBluescript SK(-); Site.1: EcoRI; Site.2:  
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were  
taken from tubers. The tubers were incubated at 26C in the  
dark for 2-3 weeks prior to sprouting. The eyes were  
frozen in liquid nitrogen immediately upon removal from  
tubers."  
BASE COUNT 220 a 149 c 181 g 218 t  
ORIGIN

alignment\_scores:  
Quality: 1143.00 Length: 231  
Ratio: 4.991 Gaps: 0  
Percent Similarity: 99.134 Percent Identity: 97.403  
alignment\_block:  
US-09-591-466c-2 x BG594788 ..  
Align seg 1/1 to: BG594788 from: 1 to: 768  
1 MetArgGLyAsnLysPheCysPheAspLeuArgTyrLeuLeuValValAl 17  
|||||  
74 ATGAGAGGGAACAAAGTTTTCCTTTGATTTACGGTACCTCTCTCGTGGC 123  
|||||  
17 aLaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34  
|||||  
124 TGCTCTCGCTTCATCTACATACAGATCGGCTTTTCGCGACACAGTCAG 173  
|||||  
34 lutyValaspargLeuAlaAlaIleGluAlaGluasnHlsCysThr 50  
|||||  
174 AATATGTAGACCCCTTGTGTCGAATTCGAACGACGAAATCATTTGACA 223  
|||||  
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlyArgVa 67

```

|||||
224 AGTCACAGCAGATGCTTATTGACAAAGATTAGCCACGACCAAGGAAGAGT 273
|||||
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGlnCysArgGlnL 84
|||||
274 AGTAGCTCTTGAGACAAATGACCGTCAGGACGAGGAGTCCGCCAAT 323
|||||
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
324 TAAGGCTCTTGTTCCAGGATCTTGAAGTAAGGGCATATAAAAAAGTTAATC 373
|||||
101 GlyAspValClnMetProValAlaAlaValValMetAlaCysSerAr 117
|||||
374 GGAGATGTCAGATGCCAGTGGCAGCTGTAGTTATGGCTTGCAGTCG 423
|||||
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
424 TGCTGACTACCTGGAGAGGACTATTAAATCCATCTTAAATACCAACAT 473
|||||
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
474 CTGTTGCATCAAAATATCCTCTTTTCATATCCAGGATGGATCAATCCT 523
|||||
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
524 GATGTAAAGAAAGCTGCTTTGAGCTATGATCAGCTGACGTATATGCAGCA 573
|||||
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaIat 184
|||||
574 CTGGATTTTGAACCTGTGCATACCTGAAAGACCGGGGAACCTGGTGCAT 623
|||||
184 yrTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeuPheHis 200
|||||
624 ACTACAGGATTCACCTCATTAACAAGTGGGCATTTGGATCAGCTGTTTTCAC 673
|||||
201 LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAl 217
|||||
674 AAGCATAATTTAGCCGTGTTATCATACATAGAGAATGATATGGAATTCG 723
|||||
217 aAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuLeu 231
|||||
724 TGCTGATTTTGTGACTATTTTGGGGCTGAGCTACTCTCTCT 766

```

seq\_name: gb\_est1:A1486771

seq\_documentation\_block:

LOCUS A1486771 643 bp mRNA linear EST 18-MAY-2001  
 DEFINITION EST245093 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
 cLED11D20, mRNA sequence.

ACCESSION A1486771

VERSION A1486771.1 GI:4382142

KEYWORDS EST.

SOURCE Tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

REFERENCE 1 (bases 1 to 643)

Aicala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,  
 Liang, P., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman,  
 C.L., Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley,  
 S.D. and Giovannoni, J.

Generation of ESTs from tomato carpel tissue

Unpublished (1999)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>.

Location/Qualifiers

1..643

/organism="Lycopersicon esculentum"

FEATURES

source

```

/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED11D20"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage=5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/notes=vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:  

XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and  

directionally cloned cDNA in vector Lambda ZAP II with 5',  

and 3' ends located at the EcoRI and XhoI sites,  

respectively.

```

BASE COUNT 179 a 128 c 144 g 192 t  
 ORIGIN

alignment\_scores:

Quality: 1116.00 Length: 214

Ratio: 5.264 Gaps: 0

Percent Similarity: 99.065 Percent Identity: 96.729

alignment\_block:

US-09-591-466C-2 x A1486771 ..

Align seg 1/1 to: A1486771 from: 1 to: 643

99 LeuIleGlyAspValGlnMetProValAlaAlaValValMetAlaCys 115

2 TTAATCGGAAATGTGCAGATGCCAGTGGCAGCTGTAGTTGTTATGCTTG 51

115 sSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrG 132

52 CAGTCGTTCTGACTACCTGGAGAAGACTATAAAATCCATCTTAAATAACC 101

132 InThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySer 148

102 AAACATCTCTTGCATCAAAATATCCTCTTTTCATATCCAGGATGGATCG 151

149 AsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMe 165

152 AATCCTGATGTTAGAAAGCTTGCTTTGAGCTATGATCAACTGACGTATAT 201

165 tGlnHisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182

202 GCAGCAGCTTGGATTTTGAACCTGTGCATACTGAAAGACCGAGGGAACCTGG 251

182 aAlaIatTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeu 198

252 TTGCATACTACAAGATTGCACGTCATTACAAGTGGCATTTGGATCAGCTG 301

199 PheHisLysHisAsnPheSerArgValIleIleLeuGluAspMetG1 215

302 TTTCAACAGCATAATTTTAGCCGTGTTATCATACTAGAAAGATGATATGGA 351

215 utileAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuA 232

352 AATTGGCGCTGATTTTGTGACTACTTTTGAGGCTGGAGCTACTCTTTTGTG 401

232 sPAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGln 248

402 ACAGAGACAAAGTCGATTATGGCTATTTCATCTTGGNAATCACAATGGACAA 451

249 ArgGlnPheValGlnAspProAspAlaLeuTyrArgSerAspPhePhe 265

452 AGGCAGTTTCGTCACAGATCCTTATGCTCTTTACCGCTCAGACTTTTTC 501

265 oGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProL 282

502 TGGTCTTGGATGGATGCTTTTCAAAATCAACTTGGTCCGAACTATCTCCAA 551

282 ystrProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHis 298

552 AGTGGCCGAGGCTTACTGCGGATGACTGGCTGAGGCTCAAGAAATCAC 601

299 ArgGlyArgGlnPheIleArgProGluValCysArgThrTyr 312  
|||||  
602 AGAGGTGCAATTTATTCGCCGAGAAGTTTCAGAACGTTT 643

seq\_name: gb\_est2:BG592816

seq\_documentation\_block:

LOCUS BG592816 642 bp mRNA linear EST 12-APR-2001  
DEFINITION EST491494 cSTS Solanum tuberosum cDNA clone cSTS289 5' sequence,

mRNA sequence.

ACCESSION BG592816

VERSION BG592816.1 GI:13610956

KEYWORDS EST.

SOURCE potato.

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 642)

van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,

Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.

Generations of ESTs from sprouting potato eyes

Unpublished (2000)

Contact: Cathy Ronning

The Institute for Genomic Research

For clone info: please contact Research Genetics, Libraries

Division tel 1-800-711-6195, email cdna@resgen.com

Seq primer: M13F-R.

FEATURES

Location/Qualifiers

1..642

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db\_xref="taxon:4113"

/clone\_lib="cSTS"

/tissue\_type="sprouting eyes from tubers"

/dev\_stage="12-14 weeks post harvest"

/lab\_host="SOLR"

/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; Various sizes of sprouting eyes (2mm to 15mm) were

taken from tubers. The tubers were incubated at 26C in the

dark for 2-3 weeks prior to sprouting. The eyes were

frozen in liquid nitrogen immediately upon removal from

tubers."

alignment\_scores:

Quality: 1113.00 Length: 213

Ratio: 5.275 Gaps: 0

Percent Similarity: 99.061 Percent Identity: 97.183

alignment\_block:

US-09-591-466c-2 x BG592816 ..

Align seg 1/1 to: BG592816 from: 1 to: 642

99 LeuIleGlyAspValGlnMetProValAlaValValMetAlaCy 115

|||||

3 TTAATCGGAGATGTCAGATGCCAGTGCAGCTGTAGTTGTATGGCTTG 52

|||||

115 sSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrG 132

|||||

53 CAGTCGTGCTGACTACCTGGAGAGACTATTAAATCCATCTTAAATACC 102

|||||

132 InThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySer 148

|||||

103 AAACATCTGTTGCATCAAAATATCCTCTTTTCATATCCAGGATGCATCA 152

|||||

149 AsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMe 165

|||||

|||||

153 AATCCTGATGTAAGAAAGCTTGCTTTGAGCTATGATCAGCTGACGTATAT 202

|||||

165 tGlnHisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182

|||||

203 GCAGCACTTGGATTTCACCTGTGCATAGTAAAGACCAGGGGAACCTGG 252

|||||

182 aAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeu 198

|||||

253 TTGCATACCTACAGATTGCAGTCATTACAAAGTGGGCAATGGATCAGCTG 302

|||||

199 PheHisLysHisAsnPheSerArgValIleIleLeuGluAspMetGI 215

|||||

303 TTTTCAACAGCATAAATTTAGCCGCTGTTATCATCTACTAGAACATGATGGA 352

|||||

215 ulleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuA 232

|||||

353 AATTTGCTGCTGATTTTTCGACTATTTTGGAGCTGGAGCTACTCTTCTTG 402

|||||

232 sPArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGln 248

|||||

403 ACAGACAAAGTCGATTATGGCTATTCTTCTTGGAAATGACAATGGACAA 452

|||||

249 ArgGlnPheValGlnAspProAspAlaLeuTyrArgSerAspPhePr 265

|||||

453 AGGCAGTTCGTCACAGATCCTGATGCTCTTTACCGCTCAGACTTTTTC 502

|||||

265 oGlyLeuGlyTyrMetLeuSerLysSerThrTrpSerGluLeuSerProL 282

|||||

503 TGGTCTTGGATGGATGCTTTTCAAAATCAACTTGTGCGAACTATCTCAA 552

|||||

282 yStrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHis 298

|||||

553 AGTGGCCAAAGGCTTACTGGGACTGGCTAAGCTGAAAGAAATATCATC 602

|||||

299 ArgGlyArgGlnPheIleArgProGluValCysArgThr 311

|||||

603 AGAGGTGCAATTTATTCACCCAGAAAGTTTGCAAAACG 641

|||||

seq\_name: gb\_est2:BI179808

seq\_documentation\_block:

LOCUS BI179808 724 bp mRNA linear EST 09-JUL-2001

DEFINITION EST520753 cSTE Solanum tuberosum cDNA clone cSTE20K20 5' sequence,

mRNA sequence.

ACCESSION BI179808

VERSION BI179808.1 GI:14645619

KEYWORDS EST.

SOURCE potato.

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 724)

van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Cho,J.,

Chiemiango,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and

Baker,B.

Generation of ESTs from in vitro grown microtubers

Unpublished (2001)

Contact: Cathy Ronning

The Institute for Genomic Research

For clone info: please contact Research Genetics, Libraries

Division tel 1-800-711-6195, email cdna@resgen.com

Seq primer: M13F-R.

Location/Qualifiers

1..724

/organism="Solanum tuberosum"

/cultivar="Bintje"

/db\_xref="taxon:4113"

/clone="cSTE20K20"

/clone\_lib="cSTE"

/tissue\_type="axillary buds of stem explants; growing

sink-tubers"

/dev\_stage="7, 8 and 10 days"

/lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; Tissue supplied by Christian Bachem and Richard  
 Visser (Department of Plant Breeding, Wageningen  
 University, The Netherlands). The cSTA libraries will  
 attempt to capture the induction and initiation/initial  
 growth of the tuber in an in vitro system as described in  
 Bachem et al. (Plant Journal, 1996). Small microtubers  
 develop from axillary buds attached to stem explants when  
 placed on a high sucrose medium (10%). Visible  
 morphological changes occur synchronously at day five in  
 the axillary buds. The first library, cSTA (1-20) consists  
 of axillary buds harvested on days 1-3. This targets  
 those genes involved in induction of the microtubers. The  
 following libraries, cSTA (21-40) and cSTA (41-60),  
 capture genes involved in tuber initiation and outgrowth.  
 This library is noted as P3 in Tanksley lab notebooks."

BASE COUNT 215 a 143 c 171 g 195 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 1112.00 Length: 218  
 Ratio: 5.101 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-591-466C-2 x B1179808 ..  
 Align seg 1/1 to: B1179808 from: 1 to: 724

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17  
 69 ATGAGAGGGAACAAAGTTTCTGATTACGGTACCTTCTCGTGGC 118  
 17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34  
 119 TGCTCTCGCTTCATCTACATACAGATGCGGCTTTTCGGCACACAGTCAG 168  
 34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50  
 169 AATATGTAGACCGCTTGCTGCTCAATGAACGACAGAAATCATTTGACA 218  
 51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67  
 219 AGTCAGACCAAGATTGCTTATTGACAAGATTAGCCAGCAGCAAGAAAGAT 268  
 67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnI 84  
 269 AGTAGCTCTGAAGAACAAATGAAGCATCAGGACAGGAGTCCGCGCAAT 318  
 84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100  
 319 TAAGGGCTCTCTTCAGGACTCTTGAAGTAAGGGCATAAAAAGTTAATC 368  
 101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117  
 369 GGAGATGTCAGATGCCATGCCAGCTGTAGTTGTTATGGCTTGCAGTCG 418  
 117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134  
 419 TACTGACTACCTGGAGAGGACTATTAAATCCATCTTAAATAACCAACAT 468  
 134 exValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150  
 469 CTGTTGTCATCAAAATATCTCTTTTCATATCCAGGATGGATCAATCCT 518  
 151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167  
 519 GATGTAAGAAAGCTTGCTTGAGCTATGTCAGCTGACGTATATGCACCA 568  
 167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184  
 |||||||

369 CTTGGATTATGAACCTGTGCATACTGAAAGACCAGGGGAACCTGGTTGCAT 618  
 184 yrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200  
 619 ACTACAAGATTGCAGTCATTTACAAGTGGGCAATTGGATCAGCTGTTTCAC 668  
 201 LysHisAsnPheSerArgValIleIleLeuGluAspMetGluIleAl 217  
 669 AAGCATAATTTTACCGCTGTATTATCATACTACAAGATGATATGGAATTCG 718  
 217 aAla 218  
 719 TGCT 722

seq\_name: gb\_est1:A1725875

seq\_documentation\_block:

LOCUS A1725875 681 bp mRNA linear EST 11-JUN-1999  
 DEFINITION BNLGH113374 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar  
 to ALPHA-1,3-MANNOSYL-GLYCOPROTEIN, mRNA sequence.

ACCESSION A1725875  
 VERSION A1725875.1 GI:5044727  
 KEYWORDS EST.  
 SOURCE upland cotton.  
 ORGANISM Gossypium hirsutum

REFERENCE 1 (bases 1 to 681)  
 AUTHORS Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.  
 TITLE ESTs from developing cotton fiber  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Ben Burr

Biological Department  
 Brookhaven National Laboratory  
 Upton, NY 11973, USA  
 Tel: 516-344-3396  
 Fax: 516-344-3407  
 Email: burr@bnl.bnl.gov

Seq primer: f3 primer.  
 Location/Qualifiers  
 1..681  
 /organism="Gossypium hirsutum"  
 /cultivar="Acala Maxxa"  
 /db\_xref="taxon:3635"  
 /clone\_lib="Six-day Cotton fiber"  
 /tissue\_type="immature fiber"  
 /dev\_stage="Six days post anthesis"  
 /lab\_host="XLI-Blue"  
 /note="Vector: pBluescript II KS+"

BASE COUNT 202 a 129 c 151 g 197 t 2 others  
 ORIGIN

alignment\_scores:

Quality: 1020.00 Length: 226  
 Ratio: 4.766 Gaps: 0  
 Percent Similarity: 94.690 Percent Identity: 78.319

alignment\_block:

US-09-591-466C-2 x A1725875 ..

Align seg 1/1 to: A1725875 from: 1 to: 681

152 ValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLe 168  
 1 GTTAAACTAAGGCTTGAGTTATTAAGGAGCTAAGCTATATATGCACCAT 50  
 168 uAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaTyr 185  
 51 AGATTATCATCCGTGTCATACAGACCGGCTGGGAATGATCGCATACT 100  
 185 yrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHisLys 201

```
|||||
101 ACAAGATTGCCCTCACTACAAATGGCATTGGATGATTGTTCTACAAG 150
202 HisAsnPheSerArgValIleIleLeuGluAspMetGluIleAlaI 218
|||||
151 CACAATTTTGACCGAGTAATAATCTTGAAGATCATATGGAATTGCC 200
218 aAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuLeuAspArgAspL 235
|||||
201 TGATTTTTTGATTACTTGGAGCAGCTGCTGCCCTTCTCGAAGGACA 250
235 ySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPhe 251
251 AGTCAATTTATGGCTGTCTCTCATGCAATGACAATGGCAAAAGCAGTT 300
252 ValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLeuG 268
|||||
301 GTGTATGACCCATATGCATCTTATCGCTCAGATTTCTTCTGGTCTGG 350
268 yTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProL 285
|||||
351 CTGCATGCTTACTAAATCTGTATGGAATGAGCTATCACCAAAATGGCAA 400
285 yAlaTyrTrpAspTrpLeuArgLeuLysGluAsnHisArgGlyArg 301
|||||
401 AGCTTACTGGGATGACTGGTTGAGATTAAAAAATATACAAATGGTCGA 450
302 GlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHisG 318
|||||
451 CAATTCCTTCGCTCAAGTATGCAGAACATATAATTTGGTGAGCATGG 500
318 ySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysLeuA 335
|||||
501 TTCAAGCATGGGCGAGTTTTCGAAAAATACCTTGCACTTATTAAGATGA 550
335 snAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeuGlu 351
|||||
551 ATGACGTCNAGTGGACTGGAAGTCNAGGATTTGAGCTACCTAAGTGAG 600
352 AspAsnTyrValLysHisPheGlyAspLeuValLysLysLysProI 368
|||||
601 GAAAATATGCCCAATACTTTCGACACATTTTGAAGCGCGCAAAACCTGT 650
368 ehISgIyAlaAspAlaValLeuLysAla 377
: |||||
651 CCTTGGACAGATTCTGCTCTTATGGCA 678
```

seq\_name: gb\_est2:BG889872

```
seq_documentation_block:
LOCUS      BG889872                      661 bp    mRNA    linear    EST 30-MAY-2001
DEFINITION EST515723 cSTD Solanum tuberosum cDNA clone cSTD15H10 5' sequence,
            mRNA sequence.
ACCESSION  BG889872
VERSION    BG889872.1 GI:14266958
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 661)
AUTHORS   van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemiango,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE     Generations of ESTs from dormant potato tubers
JOURNAL   Unpublished (2001)
COMMENT   Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13F-R.
            Location/Qualifiers
FEATURES   source
            1. .661
```

```
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD15H10"
/tissue_lib="cSTD"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; This library targets genes expressed in dormant
tubers. This library was made from sections of dormant
tuber, avoiding the buds and epidermis. Tubers were stored
for one month post-harvest at 40C. The tuber was peeled,
well away from the surface. Then it was chopped into 1-2
mm cubes and immediately frozen in liquid nitrogen. This
library is noted as P4 in Tanksley lab notebooks."

BASE COUNT      196 a      133 c      158 g      174 t
ORIGIN

alignment_scores:
    Quality: 1012.00      Length: 202
    Ratio: 5.035      Gaps: 0
    Percent Similarity: 99.505      Percent Identity: 98.020

alignment_block:
US-09-591-466C-2 x BG889872 ..
Align seg 1/1 to: BG889872 from: 1 to: 661

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrIleLeuValAlaI 17
|||||
55 ATGAGAGGGAACAAGTTTTCGTTTGATTACGGTACCTTCTCGTCGTGGC 104
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
105 TGTCTCGCCTTCATCTACATACAGATGCGGCTTTTCGCGACACAGTCAG 154
34 lutyValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
155 AATATGTAGACCCGCTTGCTGTCGCAATTGAACGACAAAATCAATTGTACA 204
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
|||||
205 AGTCAGACCAGATTGCTTATTGACAAGATTAGCCAGCAGCAAGAGAGT 254
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
255 AGTAGCTCTTGAAGACAATGAAGCGTCAGGACCAGGAGTGCCGCAAT 304
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
305 TAAGGGCTCTGTTCAAGGATCTTGAAGTAAGGGCATAAAAAAGTTAATC 354
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
|||||
355 GGAGATGTCAGATGCCAGTGGCAGCTGTAGTTGTATGGCTTGCAAGTCG 404
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
405 TGCTCAGTACCTGGAGAGAGACTATTAAATCCATCTTAAATAACCAACAT 454
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
455 CTGTTGGCATCAAAATATCTCTTTTCATATCCAGGATGGATCAAAATCCT 504
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
505 GATGTAAAGAAAGCTTGCTTGGAGCTATGATCAGCTGACGTATATATGCA 554
167 sleuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
|||||
555 CTGTGATTTTGAACCTGTGCATCTACTGAAAGACAGCGGGAACCTGGTTGCAT 604
```

184 yrtYrLysIleAlaArGHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200  
 |||||  
 605 ACTACAGATTGCACGTGCTATTACAAGTGGCATTTGGATCAGCTGTTTCAC 654

201 LysHis 202  
 |||||

655 AAGCAT 660

seq\_name: gb\_est1:AW931180

seq\_documentation\_block:  
 LOCUS AW931180 574 bp mRNA linear EST 18-MAY-2001  
 DEFINITION EST357023 tomato fruit mature green, TAMU Lycopersicon esculentum

CDNA clone cLEF43B14 5', mRNA sequence.

ACCESSION AW931180

VERSION AW931180.1 GI:8106581

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 574)

Alcala, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F.,

Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M.

, Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.

Generation of ESTs from tomato fruit tissue

Unpublished (1999)

Contact: CUGI

Clemson University

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

Location/Qualifiers

1..574

/organism="Lycopersicon esculentum"

/db\_xref="TA496"

/clone="cLEF43B14"

/clone\_lib="tomato fruit mature green, TAMU"

/tissue\_type="fruit pericarp"

/dev\_stage="mature green (3-5 days pre-ripening)"

/lab\_host="SOLR"

/note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; cLEF - Fruit were tagged at the 1cm stage and

harvested 3-5 days prior to ripening. Fruit were cut in

half to verify the seeds were indeed 'immature' and the

seeds and locules were discarded prior to freezing the

pericarp"

BASE COUNT 154 a 116 c 130 g 173 t 1 others

ORIGIN

alignment\_scores:

Quality: 988.00 Length: 190

Ratio: 5.283 Gaps: 0

Percent Similarity: 98.421 Percent Identity: 96.842

alignment\_block:

US-09-591-466c-2 x AW931180 ..

Align seg 1/1 to: AW931180 from: 1 to: 574

104 GlnMetProValAlaValValMetAlaCysSerArgThrAspTy 120

|||||

4 CAGATGCCAGTGGCAGCTGTATTGTTATGGCTTGCAGCTCTTGACTA 53

120 rleuGluArgThrIleLysSerIleLeuLysTrpGlnThrSerValAla 137

|||||

54 CCTGGAGAGACATATAAAATCATCTTAAAAATACCAACATCTGTTCAT 103

137 erLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspValArg 153  
 |||||  
 104 CAAAATATCCTCTTTTCATATCCAGGATGGATCGAATCCTCATGTAGA 153  
 |||||  
 154 LysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLeuAspTy 170  
 |||||  
 154 AAGCTTGCCTTTGAGCTATGATCAACTGACGTATATGCAGCACTTGGATT 203  
 |||||  
 170 rGluProValHisThrGluArgProGlyGluLeuValAlaTyrTyrLysI 187  
 :|||||  
 204 TGAACCTGTGCATACCTGAAAGACCAGGGAACTGGTTGCATACCAAGA 253  
 |||||  
 187 leAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsn 203  
 |||||  
 254 TTGCAGCTCATTAACAAGTGGCATTCAGCTGTTTTCACAAGCATAAT 303  
 |||||  
 204 PheSerArgValIleIleLeuGluAspAspMetGluIleAlaAlaAspPh 220  
 |||||  
 304 TTTAGCCGTGTTATCATACTAGAGATGATAGGAAATTCGGGCTGATTT 353  
 |||||  
 220 ePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLysSerI 237  
 |||||  
 354 TTTTGACTACTTTGAGCTGGAGCTACTCTTCTTGACAGAGACAAGTCGA 403  
 |||||  
 237 leMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGln 253  
 |||||  
 404 TTATGGCTATTTCATCTTGAATGACAATGGACAAGGCGAGTTCGTCCAA 453  
 |||||  
 254 AspProAspAlaLeuTyrArgSerAspPhePheProGlyLeuGlyTrpMe 270  
 |||||  
 454 GATCCCTTATGCTCTTTACCCTCAGACTTTNTTCTCTGGCTTGGATGGAT 503  
 |||||  
 270 tLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProLysAlaT 287  
 |||||  
 504 GCCTTCAAAATCAACTTGGTCCGAACTATCTCCAAAGTGGCGAAGGCTT 553  
 |||||

seq\_name: gb\_est1:AW979500

seq\_documentation\_block:

LOCUS AW979500

DEFINITION EST310521 tomato root deficiency, Cornell University Lycopersicon

esculentum cDNA clone cLEW209 5', mRNA sequence.

ACCESSION AW979500

VERSION AW979500

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 673)

van der Hoeven, R.S., Garvin, D.F., Matern, A.L., Holt, I.E., Liang, F.,

Upton, J., Hansen, T.S., Ronning, C.M., Craven, M.B., Bowman, C.L.,

Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Giovannoni, J.J.

and Tanksley, S.D.

Generation of ESTs from tomato nutrient-deficient roots

Unpublished (1999)

Contact: CUGI

Clemson University

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

Location/Qualifiers

1..673

/organism="Lycopersicon esculentum"

/cultivar="TA496"

```
/db_xref="taxon:4081"
/clone="cLEW209"
/clone_lib="tomato root deficiency, Cornell University"
/tissue_type="roots"
/dev_stage="5-6 weeks old"
/note="vector: pBluescript SK-; Site_1: 5' EcoRI; Site_2:
3' XhoI; supplier: Tanksley; Tissue supplied by Dave
Garvin (USDA-ARS, Ithaca, NY 14850). Roots were harvested
from plants grown under the following
deficiencies/stresses: 10 mM Al, Zn, P, K, Fe, N), and
mRNA was isolated from individual treatments. Proportional
aliquots of mRNA of each treatment were mixed and used for
library construction."
```

BASE COUNT 198 a 136 c 159 g 180 t  
ORIGIN

alignment\_scores:  
Quality: 974.00 Length: 201  
Ratio: 4.918 Gaps: 0  
Percent Similarity: 98.507 Percent Identity: 95.025

alignment\_block:  
US-09-591-466C-2 x AW979500 ..

Align seg 1/1 to: AW979500 from: 1 to: 673

```
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17
|||||
70 ATGAGAGGAAACAGATTTTGGTTGATTTACGGTACCTTCTCGTCGGGC 119
17 aAlaLeuAlaPheLeuGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
120 TGCTCTCGGCTTCATCTACATACAGATCGCGCTTTTCGTGACAGTCAG 169
34 LuTyrValAspArgLeuAlaAlaLeuGluAlaGluAsnHisCysThr 50
|||||
170 AATATGACAGCCCTTCTGCTGCGATTGAAGCAGAGAAATCATTTGACA 219
51 SerGlnThrArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 67
|||||
220 AGTCAGACTAGTCTTATTGACAAAGATTAGCCAGCAGCAAGGAAGAGT 269
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
270 AGTGGCTCTTGAAGAACAAATGAACGCCAGCAGCAGGATGCCGACAA 319
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeu 100
|||||
320 TAAGGGCTCTTGTTCAGGATCTTGAAGTAAGGGCATAAAAAAGTTAATC 369
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
|||||
370 GGAATGTGCAGATGCCATGCCACTGTAGTTGTTATGGCTTGACGTCG 419
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThr 134
|||||
420 TTCTGACTACCTGGAGAAGACTATAAAATCCATCTTAAATACCAACAT 469
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
470 CTGTTGCATCAAAATATCTCTTTTCATATCCAGATGGATCGAATCCT 519
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
520 GATGTTAGAAAGCTTCTTTGAGCTATGATCAACTGACGCTATATGACGA 569
167 sLeuAspTyrGluProValHisThrClnArgProGlyGluLeuValAla 184
|||||
570 CTTGGATTTTGAACCTGTGCATCTACTGAAAGACCAAGCGGAACTGGTTC 619
184 yTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
|||||
```

```
620 ACTACAAGATTGCACGTCTATTACAAGTGGGATTCGATCAGCTGGTTTCC 669
201 Lys 201
||||
670 AAG 672
```

seq\_name: gb\_est1:AW735827

seq\_documentation\_block:  
LOCUS AW735827 659 bp mRNA linear EST 18-MAY-2001  
DEFINITION EST336595 tomato flower buds 0-3 mm, Cornell University  
LYCopersicon esculentum cDNA clone cTOA5111 5', mRNA sequence.

ACCESSION AW735827  
VERSION AW735827.1 GI:7642666  
KEYWORDS EST.  
SOURCE tomato.

ORGANISM

LYCopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
LYCopersicon.

REFERENCE 1 (bases 1 to 659)

AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang  
,F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W.,  
Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

TITLE Generation of ESTs from tomato flower tissue, 0-3 mm buds

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

source Location/Qualifiers

1..659

/organism="LYCopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone="cTOA5111"

/clone\_lib="tomato flower buds 0-3 mm, Cornell University"

/tissue\_type="flower"

/dev\_stage="0-3mm buds"

/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; supplier: Tanksley; Flower buds and flowers were

taken from greenhouse plants (4-8 wks old, TA496). They

were immediately frozen in liquid nitrogen and then

size-separated while remaining frozen."

BASE COUNT 195 a 132 c 156 g 176 t

ORIGIN

alignment\_scores:

Quality: 954.00

Ratio: 4.918

Percent Similarity: 98.980

Percent Identity: 95.408

alignment\_block:

US-09-591-466C-2 x AW735827 ..

Align seg 1/1 to: AW735827 from: 1 to: 659

```
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17
|||||
70 ATGAGAGGAAACAGTTTTGGTTGATTTACGGTACCTTCTCGTCGGGC 119
17 aAlaLeuAlaPheLeuGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
120 TGCTCTCGGCTTCATCTACATACAGATCGCGCTTTTCGTGACAGTCAG 169
34 LuTyrValAspArgLeuAlaAlaAlaLeuGluAlaGluAsnHisCysThr 50
|||||
170 AATATGACAGCCGCTTCTGCTGCGATTGAAGCAGAGAAATCATTTGACA 219
```

```
/clone_lib="wheat unstressed seedling shoot cDNA library"
/tissue_type="Etiolated shoot"
/dev_stage="Five day old seedling"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Shoots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared. A cDNA library was made, and the cDNA clones
were in vivo excised to give pBluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
```

BASE COUNT 195 a 115 c 147 g 172 t  
ORIGIN

alignment\_scores:  
Quality: 953.00 Length: 209  
Ratio: 4.938 Gaps: 0  
Percent Similarity: 92.344 Percent Identity: 81.818

alignment\_block:  
US-09-591-466C-2 x BE426174 ..

Align seg 1/1 to: BE426174 from: 1 to: 629

135 ValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnProAs 151  
|||||  
2 GTTGCTTCAAGATTTCCTCACTATTATATACACAGGATGGAACAATGGAGA 51  
151 pValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisL 168  
:|||||  
52 AGTAAAAAGAAAGCTTGGATTACACTCAATAATACATTTATGCGACATG 101  
168 euAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaTyr 184  
:|||||  
102 TGGATCTTGACCTGTGCGCACTGAAAGACACGAGGAGAAACCTTGCCATAT 151  
185 TyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeuPheHisL 201  
|||||  
152 TACAAGATAGCTAACCACTATAAATGGGCTTGGATGAGCTATTCATTAA 201  
201 sHisAsnPheSerArgValIleIleLeuGluAspMetGluIleAlaA 218  
|||||  
202 GCATGATTTTCGTCGATATCATCTCTGGAAGATGACATGGAGATCGCC 251  
218 laAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAsp 234  
|||||  
252 CAGATTTCTTTGACTACTTTGAGCGTCGACGCGAAATTTACTTGACACTGAC 301  
235 LysSerIleMetAlaIleSerSerTyrAsnAspAsnGlyGlnArgGlnPh 251  
|||||  
302 AAGTCGATATGGCTGTTCTTCTTGGCAATGACAATGGCAAAAGCAGTT 351  
251 eValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLeuG 268  
|||||  
352 TGTATATGACCCCAAGGCTCTTTACCGTTGCGATTTCTTCTCGGCGCTTG 401  
268 lyTyrMetLeuSerLysSerThrTyrPheSerGluLeuSerProLysTyrPro 284  
|||||  
402 GATGATGCTTACGAGCTCAACATGGATGGAGCTGTCCACCAAGTGGCC 451  
285 LysAlaTyrTyrAspAspTyrPheArgLeuLysGluAsnHisArgGlyAr 301  
|||||  
452 AAAGCTTATTGGGATGACTGGGTAAAGGAGGTACACAGAGATCG 501  
301 gClnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHisG 318  
|||||  
502 GCAGTTTATTGCGCCAGAGATATGCAGAAACATACAACTTTTGGCGAGCATG 551

```
51 SerGlnThrArgLeuLeuLeuAspLysIleSerGlnGlnGlnArgVa 67  
|||||  
220 AGTCAGACTAGTGTGTTTACAGAGATTATCCACGACGACGAGGAGAGT 269  
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGlnCysArgGlnL 84  
|||||  
270 AGTGCTCTTCAGACAAATGAAGCGCCAGGACGAGGAGTCCGACAA 319  
84 euArgAlaLeuValGlnAspLeuSerLysGlyIleLysLysLeuL 100  
|||||  
320 TAAGGCTCTTGTTCAGAGATCTTGAAGTAAGGCGATATAAAGTTAATC 369  
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117  
|||||  
370 GGAATGTGCAGATGCCAGTGGCAGCTGTAGTTGTTATGCTTGCAGTCG 419  
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134  
|||||  
420 TTCTGACTACCTGGAGAGACTATATAATCCATCTTAAATAACCAACAT 469  
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150  
|||||  
470 CTGTTGCATCAAAATATCTCTTTTCATATCCAGGATGATGGAATPCT 519  
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167  
|||||  
520 GATGTTAGAAAGCTTGTGCTTGGATCATCACTGACGATGATATGCAGCA 569  
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaT 184  
|||||  
570 CTGCGATTTTGAACCTGTGCATCTGAAAGACACGAGGGAACCTGGTGCAT 619  
184 yTyrLysIleAlaArgHisTyrLysTyrAlaLeuAsp 196  
|||||  
620 ACTACAAGAATGCACGTCTATTACAAGTGGCATTTGGAT 657
```

seq\_name: gb\_est2:BE426174

seq\_documentation\_block:  
BE426174 629 bp mRNA linear EST 24-JUL-2000  
LOCUS WHE0329\_H01\_00125 wheat unstressed seedling shoot cDNA library  
DEFINITION Triticum aestivum cDNA clone WHE0329\_H01\_001, mRNA sequence.  
ACCESSION BE426174  
VERSION BE426174.1 GI:9424017  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
; Triticeae; Triticum.  
1 (bases 1 to 629)  
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han  
, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,  
Seaton, C.L. and Tong, J.C.  
The structure and function of the expressed portion of the wheat  
genomes  
Unpublished (2000)  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@w.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stragene SK primer.  
Location/Qualifiers  
1..629  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE0329\_H01\_001"



318 lySerSerLeuGlnPhePheLysGlnTyrLeuGluProIleLysLeu 334  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
552 GATCAAGCATGGACAATCTTTGATCAATCTTGAACCAATCAATTA 601  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
335 AsnAspValGlnValAspTrpLysSer 343  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
602 AATGATGCTCATATTGACTTGAATTC 628

seq\_name: gb\_est2:BE516817

seq\_documentation\_block:  
LOCUS BE516817 603 bp mRNA linear EST 08-AUG-2000  
DEFINITION WHE620\_D08\_H162A wheat ABA-treated embryo cDNA library Triticum aestivum CDNA clone WHE620\_D08\_H16, mRNA sequence.

ACCESSION BE516817 GI:9740847

VERSION BE516817.1

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; ; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 603)

Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C., Verhey ,S.D. and Walker-Simmons,M.K.

The structure and function of the expressed portion of the wheat genomes - ABA-treated embryo library

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: andersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20

Seq primer: Clontech Matchmaker 3' AD primer.

FEATURES

source

1..603  
/organism="Triticum aestivum"  
/cultivar="Brevor (soft, white, winter, common wheat)"  
/db\_xref="taxon:4565"  
/clone="WHE620\_D08\_H16"  
/clone\_lib="wheat ABA-treated embryo cDNA library"  
/tissue\_type="Seed embryo"  
/dev\_stage="Mature dormant"  
/lab\_host="E. coli DH12S"  
/note="Vector: pCAD10; Site\_1: EcoRI; Site\_2: XhoI;  
Embryos were cut from mature, dormant seeds and imbibed in 25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7, for 12 hr at 22 C. The tissue, total RNA, and poly(A) RNA were prepared by Steven Verhey in M.K. Walker-Simmons's lab (USDA-ARS, Washington State Univ., Pullman, Washington 99164-6420. A cDNA library was made by Clontech using a combination of random and oligo dt primers. Library was plated and archived by Russell Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 147 a 174 c 120 g 162 t

alignment\_scores:

Quality: 859.00 Length: 199

Ratio: 4.853 Gaps: 0

Percent Similarity: 88.945 Percent Identity: 77.387

alignment\_block:

US-09-591-466c-2 x BE516817/rev ..

Align seg 1/1 to reverse of: BE516817 from: 1 to: 603  
231 LeuAspArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnG1 247  
||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
602 CTTGACACTGACAAGACAATAATGGCTGTTTCATCTTGAATGATAATGG 553  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
247 yGlnArgGlnPheValGlnAspProAspAlaLeuTyrArgSerAspPheP 264  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
552 GCAAAAGCAGTTTCGTTTATGACCCCAAGGCTCTTTACCGTTCGGAGTTCT 503  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
264 heProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSer 280  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
502 TTCCGGGGCTTGGATGGATGCTAACGAACTCAACATGGATGGAGCTGTGG 453  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
281 ProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAs 297  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
452 CCAAAAGTGCCCAAGCTTATTGGGATGACTGGGTGAGGCTAAAGGAGGT 403  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
297 nhisArgGlyArgGlnPheIleArgProGluValCysArgThrTyrAsnP 314  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
402 ACACAGAGATCGGCAGTTTATTTCGGCCAGAGTATGACGAGACATACAAC 353  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
314 heGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGlu 330  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
352 TTGGCGAGCATGATCAAGCATGGGACAATTCTTTCATCAACTACTTGAAA 303  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
331 ProIleLysLeuAsnAspValGlnValAspTrpLysSerMetAspLeuSe 347  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
302 CCTATCAAGTTAAATGATGCTCATATTGACTGGAATTCGAGGACCTGAG 253  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
347 rTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLeuValLysL 364  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
252 CTACCTCAAGGAGGACAGTGTTCACCAAAATTCGGGAAAGACGTGGCTA 203  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
364 ysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIle 380  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
202 GGGCCACACCTGTGATCGATCGATGCTTTGTTGAAGCCCAATCTG 153  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
381 AspGlyAspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspI1 397  
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
152 GATGTGGACGTAAAGGATTTCAGTATGACAATCAGGCGCACTTCGAGCGTAT 103  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
397 eAlaArgGlnPheGlyIlePheGluGluTrpLysAspGlyValProArgA 414  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
102 AGCTCGTCAGTTTGGAAATATTGAAAGAGTGGAAAGGACGCTGTCCACGGG 53  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
414 laAlaTyrLysGlyIleValValPheArgPheGlnThrSerArgArg 429  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
52 CGGCTTACAAAGGCGTGGTGGTTCCGGTACAAAGAGGTCGACGCGG 6

seq\_name: gb\_est2:BG409413

seq\_documentation\_block:

LOCUS BG409413

DEFINITION 00785 leafy spurge Lambda HybridZAP 2.1 two-hybrid vector cDNA library Euphorbia esula cDNA clone 24AD 5' similar to glycyitransferase like protein, mRNA sequence.

ACCESSION BG409413

VERSION BG409413.1

KEYWORDS EST.

SOURCE leafy spurge.

ORGANISM Euphorbia esula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.

REFERENCE 1 (bases 1 to 591)

Anderson,J.V. and Horvath,D.P.

Identification of mRNAs expressed in underground adventitious buds of Euphorbia esula (leafy spurge)

Unpublished (2000)

Contact: Anderson JV

Plants Science Research

USDA/ARS, Biosciences Research Lab  
1605 Albrecht Blvd., PO Box 5674, Fargo, ND 58105, USA  
Tel: 701 239 1263  
Fax: 701 239 1252  
Email: andersj@fargo.ars.usda.gov  
Seq primer: PAD5

FEATURES  
source  
1..591  
Location/Qualifiers  
/organism="Euphorbia esula"  
/db\_xref="taxon:3993"  
/clone="24AD"  
/clone\_lib="leafy spurge Lambda HybriZAP 2.1 two-hybrid  
vector cDNA library"  
/tissue\_type="underground adventitious buds"  
/dev\_stage="3-day induced (decapitated)"  
BASE COUNT 189 a 96 c 141 g 163 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 832.00 Length: 184  
Ratio: 4.727 Gaps: 1  
Percent Similarity: 95.652 Percent Identity: 79.348  
alignment\_block:  
US-09-591-466C-2 x BG09413 ..  
Align seg 1/1 to: BG09413 from: 1 to: 591

```

262 AspPhePheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerG1 278
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
8  GAGTGTTCGGGACTTGGATGGATGGCTTACTAGATCAACCTGGGATGA 57

278 uLeuSerProLysTrpProLysAlaTyTrpAspAspTrpLeuArgLeuL 295
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
58  ACTATCCCTTAATGGCTTAAGCTTACTGGATGACTGGCTCAGGCTAA 107

295 ySGLuAsnHisArgGlyArgGlnPheLeuArgProGluValCysArgThr 311
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
108  AAGAAATCATAAAGGCGCCCAATTTATTCGTCCARAAATCTGCAGGACA 157

312 TyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnTy 328
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
158  TATACTTTGGTGACATGGTTCCAGTTTGGGCGAGTTTTTCAACACAGTA 207

328 rLeuGluProLysLysLeuAsnAspValGlnValAspTrpLysSerMeta 345
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
208  TCTTCAGGCCAATTAATAATCAATGATATTCAGTCAATTTGGAATCAATGG 257

345 sPLeuSerTyTrpLeuLeuGluAspAsnTyTrpValLysHisPheGlyAspLeu 361
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
258  ATTTGAGCTACTGATGGAGGACAGATATCCAAATATTTTCAGTGACATT 307

362 ValLysLysAlaLysProLysHisGlyAlaAspAlaValLeuLysAlaPh 378
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
308  TTAAGACAAAGCTAAAGCTATCAGTGGAAATGATGTGTTCTTAAGGCATA 357

378 eAsnIleAspGlyAspValArgIleGlnTyTrpArgAspClnLeuAspPheG 395
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
358  TAACATAGAAGGTGATGGCGTATTCAGTACAAAGACCAATTAATATTG 407

395 luAspIleAlaArgGlnPheGlyIlePheGluGluTrpLysAspClyVal 411
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
408  AAGCATTTGCAGCCAGCTTTGGCATTTTGAAGAAATGAAGGATGGGAATA 457

412 ProArgAlaAlaTyTrpLys.GlyIleValValPheArgPheGlnThrSerA 428
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
458  CCAAGGACAGCATATAAGGGGAATAGTAGTCTTCGGGTATCAAAACACAAA 507

428 rGArgValPheLeuValSerProAspSerLeuArgGlnLeuGlyValGlu 444
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
508  GACGTATATCTCTGTCGGGCGGATTTCACTGAACACAGCTTGGAAATCGAA 557

```

seq\_name: gb\_est2:BE805530

seq\_documentation\_block:

LOCUS BE805530 664 bp mRNA linear EST 06-DEC-2001  
DEFINITION ss47a01.y1 Gm-cl061 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-cl061-1825 5' similar to TR:Q9ST97 Q9ST97 ALPHA-1  
3-MANNOSYL-GLYCOPROTEIN BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE  
; mRNA sequence.

ACCESSION BE805530  
VERSION BE805530.1 GI:10236642  
KEYWORDS EST.

SOURCE soybean.  
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE 1 (bases 1 to 664)

AUTHORS

Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna  
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
R., Waterston,R. and Wilson,R.  
Public Soybean EST Project

TITLE

JOURNAL

COMMENT

Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130  
South Memorial Parkway Huntville, AL 35801 For further information  
call: (800)-533-4363 or contact via email: ccu@resgen.com

Insert Length: 1629 Std Error: 0.00

High quality sequence stop: 419.

FEATURES

source

1..664  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl061-1825"  
/clone\_lib="Gm-cl061"  
/tissue\_type="mature flowers of field grown plants"  
/lab\_host="DH10B"  
/note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:  
XhoI; The cDNA library was constructed from mRNA isolated  
from mature flowers of field grown plants for the cultivar  
Raiden. Complementary DNA was synthesized from mRNA using  
a primer consisting of a poly(dT) sequence with a XhoI  
restriction site. EcoRI adapters were ligated to the  
blunt-ended cDNA fragments followed by XhoI digestion. The  
cDNA fragments were directionally cloned into the  
EcoRI-XhoI restriction site of the pBluescript vector. The  
ligated cDNA fragments were transformed into DH10B host  
cells (GibcoBRL). This library was constructed in the  
laboratory of Dr. Randy Shoemaker."

BASE COUNT 214 a 121 c 138 g 190 t 1 others  
ORIGIN

alignment\_scores:

Quality: 825.00 Length: 221  
Ratio: 4.209 Gaps: 0  
Percent Similarity: 88.688 Percent Identity: 70.588

alignment\_block:

US-09-591-466C-2 x BE805530 ..

Align seg 1/1 to: BE805530 from: 1 to: 664

44 GluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuLeaspLys11 60



293 rgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluValCys 309  
 |||||  
 252 GACTTAAGAGAAATCATAAAGGACACAGTATTATCCGCCCAAGTATGC 301  
 |||||  
 310 ArgThrTyraSnpPheGlyGluHisGly.SerSerLeuGlyGlnPhePheL 326  
 |||||  
 302 AGAACATATAATTTT.....GGGTCTAGTTGGTGGACAGTTTCTTA 342  
 |||||  
 326 ySglnTyrrLeuGluProIleLysLeuAsnAspValcInValAsprrPrLys 342  
 |||||  
 343 AGCAATATCTTGAGCAATCAAGCTGAATGATGTCAAGGTTGATTGGAAA 392  
 |||||  
 343 SerMetAspLeuSerTyrrLeuLeuGluAspAsnTyrrValLysHisPheGl 359  
 |||||  
 393 TTAATGGATCTGAGCTATTACTGGAGGATAAATATCTATGCACCTTGC 442  
 |||||  
 359 yAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLeuL 376  
 |||||  
 443 GAATGTTATTAAAGAACTACACCTGTCTATGAGCTGACATGGTTCTAA 492  
 |||||  
 376 ySaLapheAsnIle 380  
 |||||  
 493 AGCATCTTAATATA 506  
 |||||

seq\_name: gb\_htc:AK004760

seq\_documentation\_block:

LOCUS AK004760 2681 bp mRNA linear HTC 19-JAN-2002  
 DEFINITION Mus musculus adult male lung cDNA, RIKEN full-length enriched  
 library, clone:1200014F20:mannoside acetylglucosaminyltransferase  
 1, full insert sequence.

ACCESSION AK004760

VERSION AK004760.1 GI:12836180

KEYWORDS HTC: CAP trapper.

SOURCE Mus musculus (Strain:C57BL/6J) adult male lung cDNA to mRNA,  
 clone\_lib:RIKEN full-length enriched mouse cDNA library

clone:1200014F20.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (sites)

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2 (sites)

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3 (sites)

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
 Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,  
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishibe,T., Harada,A.,  
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
 Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,  
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,  
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4 (sites)

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

# REFERENCE

## AUTHORS

5 (bases 1 to 2681)  
 Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,  
 Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,  
 Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,I.,  
 Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,T.,  
 Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,  
 Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,  
 Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,  
 Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,  
 Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,  
 Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,  
 Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
 Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamana,K., I.,  
 Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and  
 Hayashizaki,Y.

## Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
 URL:http://genome-gsc.riken.go.jp/, Tel:81-45-503-9222,  
 Fax:81-45-503-9216)

Please visit our web site (<http://genome-gsc.riken.go.jp/>) for  
 further details.

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5'-GAGAGAGAGCGCGCACTCGAGTCTTTTCTTTTCTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. Second  
 strand cDNA was prepared with the primer adapter of sequence[5'  
 GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATTAACCCGCCCCC 3']. cDNA was  
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
 XhoI. Host: SOLR.

## FEATURES

### source

1..2681  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGD:MGI:1904351"  
 /db\_xref="taxon:10090"  
 /clone="1200014F20"  
 /sex="male"  
 /tissue\_type="lung"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"

### gene

351..1694  
 /gene="Mgat1"  
 351..1694  
 /gene="Mgat1"  
 /note="data source:MGD, source key:MGI:96973,  
 evidence:ISS

### CDS

mannoside acetylglucosaminyltransferase 1  
 putative"

/codon\_start=1

/protein\_id="BAB23541.1"

/db\_xref="GI:12836181"

/translation="MLKKQTAGLWGAIFVGNALLLFFWTRPAPGRPLPSDSALG  
 DDPASLPREVHLAEADAELERQGLLOQIKEYALWRBWRVPTVAPAPRPVPT  
 PSHQIFPILVACDSTVRCLDLXHYRSPARFPPIIVSDCCHEETAQVIAVSGTA  
 VTHRQPDLSNIAVQPDHRKFGQYIKIARHYRWALGQIFNFKFPAVAVDDLEVAP  
 DFFEYFQATYPLLRTPDLSKCSAMNDNGKQMDSSKPELLYRDFDFPGGLLLAD  
 LWALEPKWPKAFWDDWRMRPEQRKACIRPEISRTMTFGKGVSHQELGDFQHLKFI  
 KLNQGFVPTOLDLSYLQOEAYDRDLAQVYGAPOLOVEKVRTNDQELGEBRVQYTS  
 RDSFKAFKALGVMDLLKSGVPRAGYRGIVTFQFGRGRVHLAPQTWTGIDPSNN"

### polyA\_signal

2658..2663  
 /note="putative"

### polyA\_site

2681  
 /note="putative"

BASE COUNT 529 a 736 c 776 g 640 t

## ORIGIN

## alignment\_scores:

Quality: 732.00 Length: 349  
Ratio: 2.905 Gaps: 11  
Percent Similarity: 72.206 Percent Identity: 42.980

## alignment\_block:

US-09-591-466c-2 x AK004760 ..

Align seg 1/1 to: AK004760 from: 1 to: 2681

```
103 ValGlnMetProValAlaAlaValValMetAlaCysSerArgThrAs 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
666 GTCAGATCCC.....ATCCCTGGTCATTGCTGTGACCCGACGAC 706

119 pTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValA 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
707 T...GTCGCGCGTGTCTGGATAAGTTGTTGCACATATCGGCCCTCA...G 750

136 laSerLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspVal 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
751 CTGAGCGTTTCCCCATTATGTGCTGAGCAGGCTGGGCATGAAGAGACA 800

153 ArgLysLeuAlaLeuSerTyrGly...GlnLeuThrTyrMetGlnHisLe 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
801 GCACAGGTCAATTGCTTCCTATGGCACTGCAGTCACACATCGCGCAGCC 850

168 uAsp.....TyrGluProValHisThrGluArgProGlyG 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
851 GGACCTGAGTAACATTGCGGTGCGGACGACCAACCGCAAGTCCAGGGT. 899

180 luLeuValAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAsp 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
900 .....TACTACAAGATTGCCAGGCACCTACCGCTGGGCACCTAGCC 938

197 GlnLeuPheHisLysHisAsnPheSerArgValIleIleLeuGluAspAs 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
939 CAGATCTTCAACAAGTTCAAGTTCCCGCGCGCTGTGTGTAGTGGAGGATGA 988

213 pMetGluIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrL 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
989 TCTGGAAGTGGCACCAAGACTCTTTGAGTACTTCCAGGCCACCTACCCAC 1038

230 euLeuAspArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsn 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1039 TGTGAGAACAGACCCCTCCCTTGTGTGTGTCTGCTTGAATGACAAAT 1088

247 GlyGlnArgGlnPheValGlnAsp.....ProAspAlaLeuTyrArgSe 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1089 GGCAAGGAGCAGATGTAGACTCAAGCAAACTGAGCTGCTCTATCGAAC 1138

261 rAspPhePheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerG 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1139 AGACTTTTTCCTGGCCTTGGATGGCTGTGTGGCTGATCTGTGGGCAG 1188

278 luLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeu 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1189 AGCTGGAGCCCAAGTGGCCCAAGGCCCTTTTGGACACACTGGATGCCCGA 1238

295 LysGluAsnHisArgGlyArgGlnPheIleArgProGluValCysArgTh 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1239 CCTGAGCAGCGGAAGAGCGGCTTGTATTCTGCCAGAAATTTCAAGAAC 1288

311 rTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnT 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1289 TATGACCTTTTGGTCCGACGGGTGTGAGCCACGGGCAGTCTTTTGACCAGC 1338

328 yrLeuGluProIleLysLeuAsnAspValGlnValAspTrpLysSerMet 344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1339 ATCTTAAGTTTCAAGCTGAACACAGCAGTTCCTCCCTTACCACGATTG 1388

345 AspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLe 361
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1389 GACCTGTGCTACCTGCAGCAGGAGGCGCTATGACCGGACTTC..... 1430

361 uValLysLysAlaLysProIleHisGlyAlaAsp.....AlaValLeuL 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1431 .....CTCGCCAGGTCTATGTGTGCCCCCAGCTACAGGTGGAGA 1470

376 ysAlaPheAsnIleAsp.....GlyAspValArgIleGlnTyr 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1471 AAGTAAGGACCAATGATCAGAAGGAGCTGGGGAGGTGCGGTACAGTAC 1520

389 ArgAspGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheG 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1521 ACTAGCAGACAGACTTCAAGGCTTTTGCCTAAGGCCCTGGGTGTCATGGA 1570

405 uGluTrpLysAspGlyValProArgAlaAlaTyrLysGlyIleValValP 422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1571 TGACCTCAAGTCTGGTGTCCCAAGGCTGGCTACCGTGGCATTGTCACTT 1620

422 heArgPheGlnThrSerArgValPheLeuValSerProAspSer 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1621 TCCAGTTTCGG...GGCCGCGTGTGCACCTGGCACCCCCACAGACA 1664
```



```

94 ..... 94
594 CTTCAAGTGCACAGTATTATTATTCATTGCTCATTTATGCAAAATGA 643
95 ..... GlyIleL 97
644 AGGAAGCCTTCAACCTGCCGTGCAGATCGGAATGGACGACAGAGTTT 693
97 ysLysLeuIleGly..... 101
694 CAATAGTAATGGGAATTCCTACAGTGAAGAGAGAAGTTAAATCTTACCTC 743
102 ..... AspValGlnMe 105
744 ATAGAACTCTTCATTCCTTATTGTAATCTGTATCTCGAAGAGAGTT 793
105 tProValAlaAlaValValMetAlaCysSerArgThrAspTrpLeuG 122
794 GGACTGTGTATAGTATCTTTCATAGGAGAGACAGATACGATATATCTAA 843
122 luArGThrIleLysSerIleLeuLys...TyrGlnThrSerValAlaSer 137
844 ATGGTGTGTAGCAACCTCGGAGAAAGATTTTCTAAAGAAATCAGTTCT 893
138 LysTyrProLeuPheIleSerGlnAspGlySer...AsnProAspValAr 153
894 GGCTTGGTGAATAATATCACCTCCTGAAAGCTATTATCTCACCTGAC 943
153 gLysLeuAlaLeuSerTyrGly.....GlnLeuThrT 164
944 GAACCTAAGGACACATTTGGAGATCTTAAGAAAGAGTAAGATGGAGAA 993
164 yrMetGlnHisLeuAspTyr.....GluProValHisThrGluArGPro 178
994 CAAAGCAAAACCTAGATATTATTTCTTAATGATGATGCTCAGGAAAAA 1043
179 GlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLe 195
1044 GGC.....ACATACTAC..... 1055
195 uAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleLeuGluA 212
1056 .....ATCCAGCTTGAAG 1068
212 sPaspMetGluIleAlaAlaAspPheAspTyrPheGluAlaGlyAla 228
1069 ATGATATTATTCTCAACACAGAATTACTTTAACACCATAAAGAATTTGCA 1118
229 ThrLeuLeuAspArgAspLys...SerIleMetAlaIleSerSerTrpAs 244
1119 CTTCAACTTTCTTCGAGGAATGGATGACTTGAGTTCTCCAGCTGGG 1168
244 nAspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeuTyrArgS 261
1169 ATTCAATTGGT.....AAAATGTTCAAGCACCTGACCCACTCTCGATTG 1212
261 erAspPhePhe.....ProGlyLeuGlyTrpMetLeu 271
1213 TCGAATTCATATTATGTTCTATAAGGAGAAGCCCATCGACTGGCTCTG 1262
272 SerLysSerThrTrpSerGluLeu...SerProLysTrpProLysAlaTy 287
1263 GACCATATTCTGTGGTCAAACTCTGCAACCCGGGAAAAGATGCAAAACA 1312
287 rTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGlyArgGlnPheI 304
1313 CTGTGAT.....CGACAGAAGGCAAAATCTCGAATTCGT.....T 1347
304 leArgProGluValCysArgThrTyrAsnPheGlyGluHisClySerSer 320
1348 TCAGACCGTCCCTTTCCAA.....CACGTTGGTCTGCATTCCTTCACTC 1391

```

```

321 LeuGlyGlnPhe.....PheLysGlnTyrLeuGluProIleLy 333
1392 ACAGGAAAAATTTCAGAAACTCAGGATAAAGATTACATGAACCATTA 1441
333 sLeuAsn.....AspValGlnValAspTrpLys. 342
1442 GCTCAAAATCCATGTAAACCCCTCGACAGAGTATCTACTTCTTGAAG 1491
343 .....SerMetAspLysSerTyrLeuLeuGluAsp..... 352
1492 TCTACCAAGGTTCATACACTGGAGAAACATTACATGGTGAGGACTTCTTC 1541
353 .....AsnTyrValLysHisPheGlyAspLe 361
1542 TGGGCTATAACCCAGTAGCTGGAGACTACATC.....CT 1576
361 uValLysLysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaP 378
1577 ATTAAATTCGACNAGCCAGTCAATGTGGAAGTTATTTGTTCCATAGTG 1626
378 heAsnIleAsp.....GlyAspValArgIleGln..... 387
1627 GCAACAGGATCATCCAGGGGATATTCTGCTCAACACACACGGTGGAGTT 1676
388 .....TyrArgAspGlnLeuAsp..... 393
1677 CTGGCTTTGAAGAGTGAAGTTTGGACATCAGCAAAAGAACCAAGACAA 1726
394 ....PheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrp...L 408
1727 ACGATTAGAAGATGGCTATTTCAGAATAGGAAATTTGAAACGGTGTG 1776
408 ysAspGlyValProArgAlaAlaTyrLysGlyIleValValPheArg.Ph 424
1777 CGAAGGGATGTGGATCCACGCCCTAAACCCCATTTTCGGCTTCCGACTT 1826
424 eGlnThrSerArgArgValPheLeuValSerProAspSerLeuArgGlnL 441
1827 TCAGTTATTC...AGAAATCTGCTGTTGGGCCATCTTAATGAGATCCA 1873
441 eu 441
1874 TA 1875

```

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.US-09-415-522-5

seq\_documentation\_block:

```

; Sequence 5, Application US/09415522A
; Patent No. 6291660
; GENERAL INFORMATION:
; APPLICANT: Gaffney, Thomas
; APPLICANT: Wendland, Juergen
; APPLICANT: Philippsen, Peter
; TITLE OF INVENTION: No. 6291660el Fungal Genes Required For No. 6291660mal Growth
; FILE REFERENCE: CGC2046
; CURRENT APPLICATION NUMBER: US/09/415,522A
; CURRENT FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: patentin ver. 2.0
; SEQ ID NO 5
; LENGTH: 6216
; TYPE: DNA
; ORGANISM: Ashbya gossypii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(6216)
US-09-415-522-5

```

alignment\_scores:

Quality: 105.00 Length: 389  
Ratio: 0.640 Gaps: 18

...

TELEPHONE: (202) 628-5197

REFERENCE: 161C-879 (707)



218	aAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspL	235
1291		
235	ysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPhe	251
1298	AAATCACACTAAATGCCACTAGCACACATCATCTACT	1338
252	ValGlnAspProAspAlaLeuTyrArgSerAspPheProClyLeuG	268

[illegible]

```

seq_documentation_block:
: Sequence 5, Application US/08840236
: Patent No. 5922578
:
: GENERAL INFORMATION:
:
: APPLICANT: MARUTA, Kazuhiko
: APPLICANT: KUBOTA, Michio
: APPLICANT: SUGIMOTO, Toshiyuki
: TITLE OF INVENTION: RECOMBINANT
: TITLE OF INVENTION: FORMS NON-REDUCED
: TITLE OF INVENTION: SACCHARIDE
: NUMBER OF SEQUENCES: 19

```



[illegible]

```
285 ysAlaTyrTrpAspTrpLeuArgLeuLysGluAsnHisArgGlyArg 301
|| :|:|:|:|
1394 AAAATAAGTCGAGGAATGG.....CAT 1416
302 GlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHisL 318
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1417 AGTATCATAAATCCAAAGGTATCAGAAATGATGATATAGATATATCA 1466
318 ySerSerLeuGlyGlnPhePhe..... 325
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1467 GCTTTAGTGGCAAGTTTATGAGGGATTCTCTAATGATTTTAAAGGAGA 1516
326 .....LysGlnTyr...LeuGluProIleLysLeuAsnAspValGlnVal 339
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1517 GAATAAAGCAACATATGATAAAAGTGTACAGAGCTAAGATAAATACC 1566
340 AspTrpLysSerMetAspLeuSerTyr.....LeuLe 350
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1567 TCATGGAGAAATCAAAATAAGAAATATGAAATAGATTAATGGAATTAGT 1616
350 uGluAspAsnTyrVal...LysHisPheGlyAspLeuValLysLysAlaL 366
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1617 GGAAGAAACTTTTACCAATAAGGATTC..... 1644
366 ySProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGly 382
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1645 .....ATTAAAGTTTC..... 1656
383 AspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaAr 399
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1657 .....ATGAATTTGAAAGTAAGATAAG 1679
399 gGlnPheGlyIlePheGluGlu.....TrpL 408
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1680 AAGGATAGGAGTATGATTAAGAGCTTATCCTTGTCGATTAAATAATATGT 1729
408 ySAspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPhe 424
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1730 CAGCCGGTATACCT...GNTTTTATCAGGGAACAGAAATATGGCGATAT 1776
425 GlnThrSerArgArgValPheLeuValSerProAspSer 437
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1777 .....TTACTTACAGATCCAGATAAC 1797
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:us-08-505-448A-5
```

```
seq_documentation_block:
; Sequence 5, Application US/08505448A
; Patent No. 5976856
; GENERAL INFORMATION:
; APPLICANT: MARUTA, Kazuhiko
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME WHICH
; TITLE OF INVENTION: FORMS NON-REDUCING SACCHARIDE FROM REDUCING AMYLACEOUS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W. Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/505,448A
; FILING DATE: 21-JUL-1994
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 190183/1994
; FILING DATE: 21-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 189706/1995
; FILING DATE: 04-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: MARUTA=3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2160 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-505-448A-5

alignment_scores:
Quality: 102.50 Length: 463
Ratio: 0.551 Gaps: 21
Percent Similarity: 40.173 Percent Identity: 17.495

alignment_block:
US-09-591-466C-2 x US-08-505-448A-5 ..
Align seg 1/1 to: US-08-505-448A-5 from: 1 to: 2160
21 PheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerGluTyrValas 37
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
826 TTCTTAAATTTACTCCAACCTTACTGTTTAAATTTTAAATCAAGAGATAATGGA 875
37 pargLeuAlaAlaIleGluAlaGluAsnHisCysThrSerGlnThrA 54
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
876 CAGTATATATGAGAAATTTTCACAGCGGAGAAAATATCTATAAGTGAAAGT. 924
54 rgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgValValAlaLeu 70
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
925 .....ATAAGAAAATAAAAGCGCAA..... 945
71 GluGluGlnMetLysHisGlnAspGlnGluCysArgGlnLeuArgAlaLe 87
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
946 .....AT..... 947
87 uValGlnAspLeuGluSerLysGlyIleLysLysLeuIleGly..... 101
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
948 AATTGATGAGCTATTTAGTTAGTTATGAAAGTTAAAGATTAGCATCACAAC 997
102 .....AspValGln 104
998 GAATTAGTACCATATATGAGAGATTACCTTTCTTGTATAGATGTGTAC 1047
105 MetProValAlaAlaValValMetAlaCysSerArgThrAspTyrLe 121
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1048 AGAAGCTTATGCTAATCAGATTGTAAGAGAGTGTGATAAGCAACCAATGAGAT 1097
121 uGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaSerL 138
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1098 AGAGGAAGCAACCAAA..... 1113
138 ySyrProLeuPheIleSerGlnAspGlySerAsnProAspValArgLys 154
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1114 .....AGAAATCCAGAGCTTATACT 1134
155 LeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGl 171
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1135 AAATTACAACAATATATGCCAGCAGTATACGCTAAA.....GCTTATGA 1178
```

```

seq_name: /cgn2_6/ptodata/2/Ina/6B_COMB.seq:US-09-357-251-13

seq_documentation_block:
; Sequence 13, Application US/09357251
; Patent No. 6271441
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Schwaber, James S.
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
; FILE REFERENCE: BB-1193
; CURRENT APPLICATION NUMBER: US/09/357,251
; CURRENT FILING DATE: 1999-07-20
; EARLIER APPLICATION NUMBER: 60/093,530
; EARLIER FILING DATE: July 21, 1998
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
; LENGTH: 2143
; TYPE: DNA
; ORGANISM: Glycine max
US-09-357-251-13

alignment_scores:
      Quality: 101.00      Length: 478
      Ratio: 0.505      Gaps: 22
Percent Similarity: 41.841      Percent Identity: 19.665

alignment_block:
US-09-591-466C-2 x US-09-357-251-13      ..

Align seg 1/1 to: US-09-357-251-13 from: 1 to: 2143

39  LeuAlaAlaLalLeGluAlaGluAsnHisCysThrSerGlnThrArgLeu 55
    ||||: ||||:||||||:||||||:||||||:||||||:||||||:||||||:
42  C|CTCCCAAGCAATGGAAGTTCCTTCGGAAGCGCGTCT...ACCGGCAT 88
    ||: ||||:||||||:||||||:||||||:||||||:||||||:||||||:
55  uLeuIlleAspLysIleSerGlnGlnGlnGlyArgValValAlaLeuGluG 72
    : ||: ||||:||||||:||||||:||||||:||||||:||||||:||||||:
89  CGCGCGCGAAACCATTAGCAAAAT.....GCGCTGAAGC 123
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:
72  luGlnMetLysHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuVal 88
    :|||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
124  GCGAAGCTCAAGAACAAACAGAGAAGAAAGAAAGCAAG..... 167
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
89  GlnAspLeuGluSerLysGlyIleLysLysLeuIleGlyAspValcInMe 105
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
168  .....GAGGAGGACAGGCCAAAAG..... 188
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
105  tProValAlaAlaValValMetAla..... 114
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
189  .....GCAGCTGAATGCAGAGGCTAAGGATAACAAATCTGCACCTG 231
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
115  .....CysSerArgThrAspTyrLeuGluArgThrIleLys 126
    :|||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
232  CTGATCAAGATGATATGGCCCACTCAATACCTTGAGAAATAGGCTA... 278
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
127  SerIleLeuLysTyrGlnThrSerValAlaSerLysTyrProLeuPheIl 143
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
279  .....AAGTATCTTGCAGTTCA 295
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
143  eSerGlnAspGlySerAsnProaspValArgLysLeuAlaLeuSer.... 158
    :||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
256  AAAGCAGAGGGGAATAACCCCTATCTCTCAAAATCTTTGCTACTATGT 345
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
159  .....TyrGlyGlnLeuThrTyrMetGlnHis 167
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
346  CTC|TTGATCAATACATCAAGGAATATGAGGTTTAAACACGGCAGCAC 395
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:
168  LeuAspTyrGluProValHisThrArgGlnArgProGlyGluLeuValAla 184
    ||||:||||||:||||||:||||||:||||||:||||||:||||||:||||||:

```

```
396 CTC.....GAGGATGCTCTGTGTCTATGGCTGGC..... 425
184 rTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHisL 201
      ::::::::::::::::::::
426 .....CGAATCATGCACA 438
201 yshisAsnPheSerArgValIleIleLeuGluAspMetGluIleAla 217
      ::::::::::::::::::::
439 AGGCACCTCTGGTTCTAACTCGTC..... 464
218 AlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAs 234
      ::::::::::::::::::::
465 .....TTTATGACCTGCACAGCTGGCTGCTCAAGGTC..... 497
234 pLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnP 251
      ::::::::::::::::::::
498 ....CAGGTATGGCTGATGCGAGTAATACAGACTTGGATGAGCGTGAAT 543
251 heValGlnAspProAspAlaLeuTyrArgSerAspPhe..... 263
      ::::::::::::::::::::
544 TTTCCAAATTCATTCTAATGTGAACGGTGGGACATAGTTGGTATCACT 593
264 ...PheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLe 279
      ::::::::::::::::::::
594 GGCTTTCCAGGC.....AAAGTAAGAAGGTTGAAC 625
279 user..... 280
626 TAGTATTTCCCAAGACTTTTGTGTGTCTCATTTGTTGCATATGA 675
281 ..ProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGlu 296
      ::::::::::::::::::::
676 TGCCAAAGGCAAAAGTCTGCTGCTGCTGCGGATAATGCAAAATTTGAAGAA 725
297 Asn..... 297
726 AATCCATGGGTACCAGGAAGTACCAGGAATCCTGCAACATATATTTTGAA 775
298 .....HisArgGlyArgGln.....PheIleArgP 306
      ::::::::::::::::::::
776 AGATCAGGAACCTAGTATCGGTAAAGCCATTGGGATTTCATGCTTAACC 825
306 roGluValCysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGly 322
      ::::::::::::::::::::
826 CAGAGGTTTCAGAAATATTT.....AAGACCCGGTCTAAATCAAT 866
323 GlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnVa 339
      ::::::::::::::::::::
867 TGTACATTAGGAGGTTCTGTGATGACCTTGATTCTTGGAGGTTGAAC 916
339 lAspTrpLysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValL 356
      ::::::::::::::::::::
917 ACCAATGATGACATCATGCTGCTGGTGGAGCTGCGCCCGTCCATTGTAA 966
356 yshisPheGlyAsp..... 360
967 CTACATCAATGATCTTACATGAGGTTTATCATGAGGATTGCTCCAGAA 1016
361 LeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLeu.... 375
      ::::::::::::::::::::
1017 CTGTATCTTAAGGAGTGTGTTGTTGGTGGACTGGATCGTGTATTATCAAT 1066
376 ....LysAlaPheAsnIleAspGly..... 382
1067 TGTAAACAATTTAGGAATGAGGGCATAGATTGACCCCAATAATCTCGAGT 1116
383 .....AspValArgIleGlnTyrArgAspGlnLeuAspPheGlu 395
      ::::::::::::::::::::
1117 TTACTACCTGTGAGTCTATATGCGTTATAAGGACTACACGACTTGATC 1166
396 AspIleAlaArgGlnPhe.....GlyIlePheGluGluTrpLysAspG 410
      ::::::::::::::::::::
1167 GATATAACAGACCAATGTTGAGTGGTATGGTTAAGGAA..... 1205
```

```
410 yValProArgAlaAlaTyrLysGlyIleValVal 421
      ::::::::::::::::::::
1206 .CTTACCAAGCAGCTATAAAATCAAGTATCATG 1238
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-117-860-23
seq_documentation_block:
; Sequence 23, Application US/09117860A
; Patent No. 6338955
; GENERAL INFORMATION:
; APPLICANT: OGURI, Suguru
; APPLICANT: MINOWA, Mari
; APPLICANT: YOSHIDA, Aruto
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: TAKEUCHI, Makoto
; TITLE OF INVENTION: NOVEL 1-4 N-ACETYLGLUCOSAMINYLTTRANSFERASE AND GENE
; FILE REFERENCE: 081356/0119
; CURRENT APPLICATION NUMBER: US/09/117, 860A
; EARLIER APPLICATION NUMBER: WO PCT/JP97/04546
; EARLIER FILING DATE: 1997-12-10
; EARLIER APPLICATION NUMBER: JP 161462/1997
; EARLIER FILING DATE: 1997-06-18
; EARLIER APPLICATION NUMBER: JP 332411/1996
; EARLIER FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 2115
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (136)..(1740)
US-09-117-860-23
alignment_scores:
      Quality: 98.50      Length: 568
      Ratio: 0.402      Gaps: 31
      Percent Similarity: 43.134      Percent Identity: 20.423
alignment_block:
US-09-591-466C-2 x US-09-117-860-23 ..
Align seq 1/1 to: US-09-117-860-23 from: 1 to: 2115
10 LeuArgTyrLeuValValAlaAlaLeuAlaPheIle.....TyrI 24
      ::::::::::::::::::::
142 CTCGCAATGGAAGTGTAGCCACTGCTTTAGCATTTATCATTCTCTTCT 191
24 eGlnMet.....ArgLeuPheA 30
      ::::::::::::::::::::
192 TACTTTGTCTGTGTATCTACTACATGGCAAAATGGAAAGAAAACTGATTG 241
30 lAthrGlnSerGlyTrpValValAspArgLeuAlaAlaIleGluAlaGlu 46
      ::::::::::::::::::::
242 CTTATCAACGAGAATTCCTT...GCTTTGAAAGAACGCTCTTCGAATAGCT 288
47 AsnHisCysThrSerGlnThrArgLeuLeuIleAspLysIleSerGlnG 63
      ::::::::::::::::::::
289 GAACACAGATATCTACAGCGCTCTTCTGATTAATACGATTGTGCAACA 338
63 nGlnGlyArgValValAla.....LeuGluGluGlnMetL 75
      ::::::::::::::::::::
339 GTTCAAGCGGTGTAGGAGCAGCAAAACAATGGAATAGGATGCGTTGATA 388
75 yshisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValGlnAsp... 90
      ::::::::::::::::::::
389 AGTTTTCAGATAATACCTAAAGCTGTAAAGAGTGTAAAGAGTTAACAAGCAAAAA 438
```

```

91 .....LeuGluSe 93
439 TCTCTCAAGTGCAAGTATTATTATCATTTCCCTCAATTTATTGAAAA 488
93 rLysGly.....
489 TGAAGGAAGTCTTCAACCTCTGTACAGATTGCAACGGGAAGACAGGAG 538
95 .....
539 TTCAATAGTCATGGGCATTTCCACAGTGAAGAGAGAAAGTTAAATCTTAC 588
96 .....ileLysLysLeuileGlyAspValG1 104
589 CTCATAGAAACTCTTCATTCCTTATTGATAACCTGTATCCTGAAGAGAA 638
104 nMetProValAlaValValMetAlaCysSerArgThrAspTyrL 121
639 GTTGGACTGTGTTATAGTCTTTCATAGGAGACAGATATTGATTATG 688
121 euGluArgThrIleLysSerIleLeuLys...TyrGlnThrSerValAla 136
689 TACATGTGTGTAGCCACCTGGAGAAGAATTTTCTAAAGAAATCAGT 738
137 SerLysTyrProLeuPheIleSerGlnAspGlySer...AsnProAspVa 152
739 TCTGGCTTGTGGGAAGTCATATCACCCCTGAAAGCTATTATCTGCATT 788
152 lArgLysLeuAlaLeuSerTyrGly.....GlnLeuT 163
789 GACAACCTAAAGAGACATTTGGAGACTCCCAAGAAAGAGTAAAGATGA 838
163 hrTyrMetGlnHisLeuAspTyr.....GluProValHisThrGluArg 177
839 GAACAAGCAAAACCTAGATTACTGTTTCTTAATGATGTATGCTCAAGAA 888
178 ProGlyCluLeuValAlaTyrLysIleAlaArgHisTyrLysTrpAl 194
889 AAGGCC.....ATATATTAC.....
194 aLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleLeuG 211
904 .....ATTCAGCTTG 913
211 luAspAspMetGluIleAlaAlaAspPheAspTyrPheGluAlaGly 227
914 AAGATGATATTATTGTCAACAAATAATTTTAAATACCATAAAAAATTTT 963
228 AlaThrLeuLeuAspArgAspLys...SerIleMetAlaIleSerSerTr 243
964 GCATTCAACTTCTCTCAGGAATGGATGATCTAGAGTTTTCACAGCT 1013
243 pAsnAspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeuTyrA 260
1014 GGCCTTCATTGGT.....AAAATGTTCAAGCGCGGATCTTACTCTGA 1057
260 rgSerAspPhePhe.....ProGlyLeuGlyTrpMet 270
1058 TTCTAGATTCAATTCATGATGTTTCAAGAGGAAACCCATGATGGCTC 1107
271 LeuSerLysSerThrTrpSerGluLeu...SerProLysTrpProLysAl 286
1108 CTGGACCATATTCTCTGGTGAAAGCTGCAACCTGAAAGAGATGCAAA 1157
286 atyTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGlyArgGlnP 303
1158 ACATTGTGAT.....AGACAGAAAGCAAAATCTGCGAATTCGC.... 1194
303 heIleArgProGluValCysArgThrTyrAsnPheGlyHisGlySer 319
1195 ..TTCAGACCTCCCTTTTCCAA.....CATGTGGTCTGCACATCATCA 1236
320 SerLeuGlyGlnPhe.....PheLysGlnTyrLeuGluProI1 332

```

```

1237 CTATCAGGAAAAATCCAAAAAATCACGGATAAAGATTATATGAACCAATT 1286
332 eLysLeuAsn.....AspValGlnValAspTrpL 342
1287 ACTTCTTAAAAATCCATGATAAACCCACCTCGCGGAGGTATCTACTTCCTTGA 1336
342 ys.....SerMetAspLeuSerTyrLeuLeuGluAsp... 352
1337 AGGTCTACCAAGGCATACGCTGCAGAAAAAATTTACATGGGAGAGGATTC 1386
353 .....AsnTyrValLysHisPheGlyAs 360
1387 TCTGGGCTATCACACCGATAGCTGGAGACTACATC..... 1422
360 pLeuValLysLysAlaLysProIle..... 368
1423 .TTGTTTAAATTTGATAAAACCAAGTCATGTAGAAAGTTATTTGTTCCATA 1471
369 .....HisGlyAlaAspAlaValLeuLysAla..... 377
1472 GCGGCAACCAAGAACATCTCTGGAGATATTCTGCTAAACACAACTGTGGAA 1521
378 .....PheAsnIleAspGly...AspValArgIleGlnTyrArgAs 390
1522 GTTTTCCTTTTAAGAGTGAAGGTTTGGAAATAGCAAAAGAAATTTGAGAATG 1571
390 pGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluT 407
1572 CAAA...CGATTAGAAGTGGCTATTTCAGAAATAGCAAAATTTGAGAATG 1618
407 rp...LysAspGlyValProArgAlaAlaTyrLysGlyIleValValPhe 422
1619 GTCTTCGCAAGGAATGGTGGATCCAAAGTCTCAATCCCATTTTCAGCCTTT 1668
423 Arg.PheGlnThrSerArgArgValPheLeuValSerProAspSerLeuA 439
1669 CGACTTTTCAGTTATTC...AGAATTCTGCTGTTTGGGCCATTTCTTAATGA 1715
439 rg 439
1716 GA 1717

```

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-443-501A-2

```

seq_documentation_block:
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; FILE REFERENCE: 30062-20031-00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA

```

[illegible]





```

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123.912A
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/040.802
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 1419
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-123-912-111

```

```

alignment_scores:
    Quality: 94.00      Length: 382
    Ratio: 0.563        Gaps: 17
    Percent Similarity: 43.717    Percent Identity: 17.801

alignment_block:
US-09-591-466c-2 x US-09-123-912-111 ..

Align seg 1/1 to: US-09-123-912-111 from: 1 to: 1419

64  GinGlyArgValAlaLeuGluGlnMetLysHisGlnAspGlnI 80
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
244 CGAGAGCGCTCAGAATAAAGCGTGAAGAAAGAGGTTTCACTCTGAAAAAGA 293

80  ucysArgGlnLeuArgAlaLeuValGlnAspLeuGluSerLysGlyIle 97
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
294 GACGAGAGCTCAGAATAAAGCGTGAAGAAAGAGGTTTCACTCTGAAAAAGA 343

97  ysLysLeuIleGlyAspValGlnMetProValAlaAlaValValMet 113
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
344 AGGCTGAAGGAAAGAGATTGAG..... 366

114 AlaCysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeu 130
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
367 .....AACAGAGCGCATCATCAACAAATCCAAAAAGTTTGTGAC 407

130 stvGlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAsp 147
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
408 TCAATAAAGCAAACTCACATAATGATTATGAACCTGAACATAACCAAC.... 453

147 lysSerAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeu 163
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454 .....AGGCTGTTTGGAGAAAAACA 474

164 TyrMet.....GlnHisLeuAspTyrGluProValHisThrGluAr 177
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
475 TACCTCTCTTCAAAAATACATTAGATTAT..... 504

177 gProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrLysTrp 194
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
505 .....GTTGAAAAATATTATCATGCAT 526

194 laLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleLeu 210
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
527 CTCTGGAACCTGTT..... 540

211 GluAspAspMetGluIleAlaAlaAspPheAspTyrPheGluAlaG 227
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
541 .....GATTTTGTAAATGCAGC 557

227 yAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSerTrpA 244
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
558 C.....GATGAAGTCGAAAGAAAGATTAAATTCCTGGG 589

244 snAspAsnGlyGlnArgGlnPheValGlnAsp.....ProAspAlaLeu 258
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
590 TTGAAAGCAAAACAATGAAAAATCAAGGACTTGTCCCGATGGCTCT 639

```

```

259 TyrArgSerAspPheProGlyLeuGlyTyrPheMetLeuSerLysSerTh 275
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
640 ATTAGTAGC..... 648

275 rTrpSerGluLeuSer.ProLysTrpProLysAlaTyrTrpAspAspTrp 291
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
649 .....TCTACCAAGCTGG.....TGCTGGTCAACATGG 676

292 LeuArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVa 308
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
677 TTTATTTTAAAGGGCAATGGACAGGGAGT.....TTAAGAAAGAAAT 720

308 lCysArgThrTyrAsnPheGly..GluHisGlySerSerLeuGly..... 322
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
721 ACTAAGGAAGAGAAATTTGGATGAATAAGACACAAAGTAAATCTGTACA 770

323 .....GlnPhePheLysGlnTyrLeuGluProIleL 333
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
771 GATGATGACACAGAGCATTCCTTTAGCTTCACTTTCTGGAGGACTTGC 820

333 ysLeuAsnAspValGlnValAspTrpLysSerMetAspLeuSer...Tyr 348
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
821 AGGCCAAATTTAGGGATTCCATATATAAAACACACCTAACCATGTTT 870

349 LeuLeuGluAspAsnTyrValLysHisPheGlyAspLeuValLysLys 365
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
871 GTGCTTCTGCCAACACATCGATGCGCTGGAGAGATAATAGATAAAAT 920

365 alysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAsp 382
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
921 AGTCTCT.....GAGAAATTTGGTAGAGTGGACTAGTCCAG 955

382 lyAspVal.....ArgIleGlnTyr 388
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
956 GCATATGGAAGAAAGGAGTGAATCTGCACCTGCCCGGTTTGAGGTG 1005

389 ArgAspGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePhe 405
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1006 GAGGACAGTTACGATCTAGAGCGGCTCGCTGCATCGGATGGGATGGCGA 1055

405 uGluTrpLysAspGlyValProArgAlaAlaTyrLysGlyIle 419
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1056 TGCTTCACTGAG.....CACAAACCCGACTACTCGGAATG 1092

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-104-068-3

seq_documentation_block:
; Sequence 3, Application US/09104068
; Patent No. 6238882
; GENERAL INFORMATION:
; APPLICANT: Kallender, Howard
; APPLICANT: Palmer, Leslie M.
; APPLICANT: Fedon, Jason C.
; APPLICANT: Lenox, Anna L.
; APPLICANT: Wang, Min
; APPLICANT: Jaworski, Deborah D.
; TITLE OF INVENTION: gidal
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,068
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051,379
; FILING DATE: 01-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-104-068-3

alignment_scores:
      Quality: 92.50      Length: 462
      Ratio: 0.465      Gaps: 19
      Percent Similarity: 43.074      Percent Identity: 17.100

alignment_block:
US-09-591-466c-2 x US-09-104-068-3 ..
Align seg 1/1 to: US-09-104-068-3 from: 1 to: 1871

```

```

23 TyrIleGlnMetArgLeuPhe..... 29
|||||.....
196 TACATCCAGATGAAGTCTCACACAGGAGGCGCCACCCCTTCGTGTC 245
30 .....AlaThrGlnSerGluTyrValAspArgLeuAlaAla 42
|||||.....
246 CCTCGTCGCGAGGCTGATGAAGAACTTACTTAAGGAAATCGCAAGA 295
42 laIleGluAlaGluAsnHisCysThrSerGlnThrArgLeuIleAsp 58
.....
296 CAGTTGAAATCAAGAAATCGACCTTCGT...CAAACCATGATTGAT 342
59 LysIleSerGlnGlnGlnArgValValAlaLeuGluGlnMetIly 75
.....
343 GAGATTTCGTGCAAGATGCAAGTTGTCGGTGTG...CGTACAGCCAC 389
75 sHisGlnAspGlnGluCysArgGlnLeu..... 84
|||||.....
390 CCATCAAGAAATATGCTGCTAAGGCTGTATTGTGACGACGGGACTGCTC 439
85 .....ArgAlaLeuValGlnAspLeuGlu...SerLysGlyIleLys 97
.....
440 TCGTGGGAAATATATCATCGGAGACCTCAAGTACTCATCAGTTCTAAC 489
98 LysLeuIleGlyAspValGlnMetProValAlaAlaValValMetAl 114
.....
490 CACAGCTTGGCTTCTATTAACTCA..... 513
114 aCysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLys 131
.....
514 .....GCTGACAATCTCAAGGAAGTGGTCTCGAATCGGTGCT 553
131 yrGlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGly 147
.....
554 TCAAGACA.....GGA 564
148 SerAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTy 164
.....
565 ACCCTCCAGCGTCAAGGCTTCTTCTATCAATTAC...GATGACCGGA 611
164 rMetGlnHisLeuSerTyrGluProValHisThrGluArgProGlyGluL 181
.....

```

```

612 AATTACCCAGGAGACCAAGTGCCTAATCAT..... 642
181 euValAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGln 197
642 ..... 642
198 LeuPheHisLysHisAsnPheSerArgValIleIleLeuGluAspAspMe 214
642 ..... 642
214 tGluIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuL 231
643 .....TTCTCATAC.....ACTTCAC 658
231 euAspArgAspLysSerIleMetAlaIleSerSerTrp.....Asn 244
659 GTGATCAGGATTATGTCAAAGATCAAGTGCCTGCTGCTGACCTATACC 708
245 AspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeuTyrArgSe 261
709 AATGGTACCAGTCATCAGATTATCCAA.....ACAACCTCCACCGTGC 752
261 rAspPhePheProGlyLeuGlyTyrMetLeuSerLysSerThrTrpSerG 278
753 GCCTATGTTTACAGGTGTG.....GTCAAGG 778
278 luLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeu 294
779 GAGTGGGCTCGTACTCGCTGATTGAAGACAAGATTGTGGCTTT 828
295 LysGluAsnHisArgGlyArgGlnPheIleArgProGluValCysArgTh 311
829 GCGGACAGGAACGTCACCAACTCTCTCTTGAGCCAGAA..... 867
311 rTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnT 328
868 .....GGACGCAATACTGAGGAAGTCTATGTTCAG 898
328 yrLeuGluProIleLysLeuAsnAspValGlnValAspTrpLysSerMet 344
899 GACTTTCAACCAAGTCTGCTGAGGATGTCAG..... 930
345 AspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLe 361
931 .....CGTGACTT 938
361 uValLysLysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlap 378
939 GGTTCATTCCATCAAAAGCTTTGGAATAATGCAGAGATGATGCGGACAGTT 988
378 heAsnIleAspGlyAspValArgIle.....GlnTyrArgAspGlnLeu 392
989 ATGTATTGAGTATGATATGTTGCTTCATCATCAGTTGCTGCGACTTTG 1038
393 AspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTyrLys... 408
1039 GAAACCAAGAAATCTCA.....GGTCTCTTCACTGCTGCTGCAGAC 1079
409 .....AspGlyValProArgAlaAlaTyrLysGlyIleVal.... 420
1080 AAATGGAACATCAGTTATGAAAGAGCTGCTGGCCAAGGGATTATCGCGG 1129
421 .....ValPheArgPheGlnThrSerArgArgValPheLeuVal 433
1130 GTATCAATGGGCTCTGAAATCCAAGTAAACCTGAGTTGATTCTAAAA 1179
434 SerProAspSerLeuArgGlnLeuGlyValIleGluAsp 445
1180 CGAAGTGCAGCTTATATCGGGGTGATGATCGACGAC 1215
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq: US-09-104-068-1
seq_documentation_block:

```



Wed Aug 14 11:48:27 2002

us-09-591-466c-2.rni

```

1077 GGTTCATCCATCAAGGTTTGGAAATGCGAGATGATCGGACAGGTT 1126
      |||   ... |||   .....:||||:   ...   :|||   .....
378  heAsnIleAspGlyAspValArgIle.....GlnTyrArgAspGlnLeu 392
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1127 ATGCTATTGAGTATGATATGCTTGGCTTCATCATGCTGCGTGGACTTTG 1176
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
393  AspGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLys... 408
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1177 GAAACCAAGAAATCTCA.....GCTCTCTCACTGCTGGTGCAGAC 1217
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
409  .....AspGlyValProAlaAlaTyrLysGlyIleVal.... 420
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1218 AATGCAACATCAGGTTATGAAGAAGCTGCTGGCCAAGGATTATCGCGG 1267
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
421  .....ValPheArgPheGlnThrSerArgArgValPheLeuVal 433
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1268 GTATCAATGCGGCTCTGAAATCCCAAGGTAAACCTCAGTTGATTCTAANA 1317
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
434  SerProAspSerLeuArgGlnLeuGlyValGluAsp 445
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1318 CGAAGTGACGGTTATATCGGGGTGATGATCGACGAC 1353
      :|: |||:   |||:   :|||:   |||   |||   |||   |||

```

seq\_name: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq:US-08-617-785-11

```

seq_documentation_block:
; Sequence 11, Application US/08617785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttnner, Irene
; APPLICANT: Knopfel, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; FILE REFERENCE: 4-19679/A/PCT
; CURRENT APPLICATION NUMBER: US/08/617,785E
; CURRENT FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: PCT/EP94/02991
; EARLIER FILING DATE: 1994-09-07
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0
; EARLIER FILING DATE: 1993-09-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2745
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2745)
; US-08-617-785-11

```

```

alignment_scores:
Quality: 91.50 Length: 270
Ratio: 0.732 Gaps: 14
Percent Similarity: 46.296 Percent Identity: 22.222

```

US-09-591-466C-2 x US-08-617-785-11 ..

Align seg 1/1 to: US-08-617-785-11 from: 1 to: 2745

```

44  GluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuAspLysII 60
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1153 GACACAGATCCAAATGACACAGGACAGAGAAATGGAAAAGATTCCAA 1202
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
60  eserGlnGlnGlnGlyArg.....ValValAlaL 70
      :|: |||:   |||:   :|||:   |||   |||   |||   |||

```

```

1203 CTATCAGCAGGAGGTAAGTCCAGTTCGTGATTGACCCAGTCTATGCTA 1252
70  euGluGluGlnMetLysHisGlnAspGlnGlu...CysArgGlnLeuArg 85
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1253 TGGTTCAGCCCTTCCACACATGAACAAGGATCTCTGTGCTGACTACCGG 1302
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
86  AlaLeuValGlnAspLeuSerLysGlyIleLysLysLysLysLysLys 102
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1303 GGTGTCTGCCAGAGATGGAGCAAGCTGGAGGCAAGAAGTTGCTGAAGTA 1352
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
102  pValGln.....MetProValAlaAlaValValMetAlaLac 115
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1353 TATAGCAANTGTTAATTCAATGTTAGTGTGGCACTCCAGTGATTTTA 1402
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
115  ysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyr 131
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1403 ACAAGAACGGGATGCACCTGGCGT.....TATGACATCTTTCAGTAC 1446
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
132  GlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspLysE 148
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1447 CAGACCACA.....AACCCAG 1463
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
148  tAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrM 165
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1464 CAACCCGGGTTACCGTCTGATC.....GGCAGTGGACAGACG 1501
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
165  etGlnHisLeuAspTyrGluProVal.His.....ThrGluar 177
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1502 AACTTCAGCTCAATATAGAAGACATGCAGTGGGTAAAGGAGTCCGAGAG 1551
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
177  gProGlyGluLeuValAlaTyrTyrLysIleAla.....A 189
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1552 ATACCCGGCTCAGTGTGCACACTACCATGTAAAGCCAGGACAGAGAA 1601
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
189  rGHisTyrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSer 205
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1602 GACACAGAAAG.....GAACCTCTTGTGCTGTGGACCTGTGAGC 1639
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
206  ArgValIleLeuLeuGluAspAspMet.....GluIleAl 217
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1640 CTGTGCGATGTTTACAGTACCAGTACCATGTTTGTAGATGACATGCCAGCATTGC 1689
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
217  aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1690 CCT.....ATGACCAGAGAGGCCCAATGAAAATCGAACCCG 1724
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
234  sPlySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1725 ATGCCAGGATATTCCTCATCATCAACTGGAGTGGCACTCCC..... 1765
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
251  PheValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLe 267
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1766 .....CCTGGGCT 1773
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
267  uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerPro.....L 282
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1774 GTATTCTCTCTCTGCGCAATGTTGGGATCATTTGCCACCATCTTTGT 1823
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
282  yStrpPro 284
      :|: |||:   |||:   :|||:   |||   |||   |||   |||
1824 CATGGCCA 1831
      :|: |||:   |||:   :|||:   |||   |||   |||   |||

```

seq\_name: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq:US-08-617-785-13

```

seq_documentation_block:
; Sequence 13, Application US/08617785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttnner, Irene
; APPLICANT: Knopfel, Thomas

```

;; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,

;; FILE REFERENCE: 4-19679/A/PCT

;; CURRENT APPLICATION NUMBER: US/08/617,785E

;; EARLIER FILING DATE: 1996-03-19

;; EARLIER APPLICATION NUMBER: PCT/EP94/02991

;; EARLIER FILING DATE: 1994-09-07

;; EARLIER APPLICATION NUMBER: EPO 9416553.7

;; EARLIER FILING DATE: 1994-08-19

;; EARLIER APPLICATION NUMBER: EPO 93810663.0

;; EARLIER FILING DATE: 1993-09-20

;; NUMBER OF SEQ ID NOS: 26

;; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 13

;; LENGTH: 2766

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: (1)..(2766)

US-08-617-785-13

alignment\_scores:

Quality: 91.50 Length: 270

Ratio: 0.732 Gaps: 14

Percent Similarity: 46.296 Percent Identity: 22.222

alignment\_block:

US-09-591-466c-2 x US-08-617-785-13 ..

Align seg 1/1 to: US-08-617-785-13 from: 1 to: 2766

```
44  GluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAspLysI1 60
      ::::::::::: |||||||:||||| ::::::::::: |||:::
1153 GACACAGATCGGAATGCGAAGTGCACAGGAGGAGAGAAATGGAAAGATTCCAA 1202
      ::::::::::: |||:::
60  eSerGlnGlnGlnGlyArg.....ValValAlaL 70
      ::::::::::: |||:::
1203 CTATGAGCAGGAGGGTAAAGTCCAGTTCGTGATTGACGCGAGTCTATGCTA 1252
      ::::::::::: |||:::
70  euGluGlnMetLysHisGlnAspGlnGlu...CysArgGlnLeuArg 85
      :: :::: |||:|||||:||||| :::: |||
1253 TGGCTCACGCCCTTCACCATGTAACAGGATCTCTGTGCTGACTACCGG 1302
      ::::: ::::: |||:|||||:|||||:
86  AlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIleGlyAs 102
      ::::: ::::: |||:|||||:|||||:
1303 GGTGCTGCCACAGATGAGCAGAGTGGAGGCAAGAGTTGCTGAAGTA 1352
      ::::: ::::: |||:|||||:|||||:
102 pValGln.....MetProValAlaAlaValValMetAlaC 115
      ::::: |||:|||||: |||:::
1353 TATACCGCAATGTTAATTTCAATGGTAGTGTGTCGACATCCAGTGTGTTA 1402
      ::::: |||:::
115 ysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyr 131
      ::::: |||::: |||::: |||:|||||:|||||
1403 ACAAGAACGGGGTCCACCTGGCGGT.....TATGACATCTTTCAGTAC 1446
      ::::: |||::: |||:::
132 GlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySe 148
      |||||::: |||::: |||::: |||:::
1447 CAGACCACA.....AACACCAG 1463
      ::::: |||:::
148 rAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrM 165
      |||||::: |||::: |||::: |||::: |||:::
1464 CAACCCGGGTACCGTCTGATC.....GGGCAGTGCAGACAGC 1501
      ::::: |||::: |||::: |||::: |||::: |||:::
165 etGlnHisLeuAspTyrGluProVal.His.....ThrGluAr 177
      ::::: |||::: |||::: |||::: |||::: |||:::
1502 AACTTCAGCTCAATATAGAAGACATGAGTGGGGTAAAGGAGTCCGAGAG 1551
      ::::: |||::: |||::: |||::: |||::: |||:::
177 gProGlyGluLeuValAlaTyrTyrLysIleAla.....A 189
      |||||::: |||::: |||::: |||::: |||::: |||:::
1552 ATACCGCGCTCAGTGTGCACATACCATGTAAGCCAGGACAGAGAAAGAA 1601
```

```
189 rgHisTyrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSer 205
      ||||| ||| ::::: |||:|||||: |||
1602 GACACAGAAAG.....GAACTCCTTGTCTGCTGGACCTGTGAGC 1639
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
206 ArgValIleIleLeuGluAspMet.....GluIleAl 217
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1640 CTTGCGATGTTACCAGTACCAGTTTGTGATGATGACATGCCAGCATTGC 1689
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
217 aAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234
      | ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1690 CCT.....ATGCCAGAGGCCCAATGAAATCGAACCCGG 1724
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
234 sLysSerIleMetAlaIleSerSerTrpAsnAspGlnArgGln 250
      || |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1725 ATGCCAGGATATTCATCATCAAACTGGAGTGGCACTCCC..... 1765
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLe 267
      || ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1766 .....CCTGGGCT 1773
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerPro.....L 282
      | ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1774 GTGATTCTCTCTCTCGCAATGTTGGGGATCATTTGCCACCATCTTTGT 1823
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
282 ysTrpPro 284
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1824 CATGGCCA 1831
```

OM of: US-09-591-466c-2 to: N\_Geneseq\_032802.\* out\_format : pfs  
 Date: Aug 13, 2002 10:47 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
 -MODEL=framet.p2n.model -DEV=xlp  
 -O=cpn2\_1/USPTO.spool/US09591466/runat\_12082002\_095313\_4300/app\_query.fasta\_1.511  
 -DB=N\_Geneseq\_032802 -QPM=fastap -SUFFIX=ring -GAPOP=12\_000  
 -GAPEXT=4\_000 -MINMATCH=0\_100 -LOOPCL=0\_000 -LOOPEXT=0\_000  
 -QGAPOP=4\_500 -QGAPEXT=0\_050 -XGAPOP=10\_000 -XGAPEXT=0\_500  
 -FGAPOP=6\_000 -FGAPEXT=7\_000 -YGAPOP=10\_000 -YGAPEXT=0\_500  
 -DELOP=6\_000 -DELEXT=7\_000 -START=1 -MATRIX=biocsum62  
 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
 -USER=US09591466 -CCN1\_1\_525 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
 Query: US-09-591-466c-2  
 Query length: 446  
 Database: N\_Geneseq\_032802.\*  
 Database sequences: 1736436  
 Database length: 858457221  
 Search time (sec): 268.160000

WARN: APPL\_compute\_dev\_space\_s: pointers violation: prPtr=5ble76c !

#### score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA78001 +	2345.00	4502.95	1.8e-242	16	
/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA78002 +	2198.00	4218.34	1.3e-226	17	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA39340 +	1776.00	3403.40	3.3e-181	16	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA350197 +	1757.50	3367.42	3.2e-179	16	
/SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:AA062625 +	735.50	1386.83	6.7e-69	25	
/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA78003 +	727.00	1388.31	5.6e-69	51	
/SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:AA25594 +	726.50	1369.75	6.0e-68	24	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA19837 +	723.50	1370.85	5.2e-68	32	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA25595 +	719.50	1353.30	5.0e-67	32	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA22693 +	669.50	1251.40	2.3e-61	22	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA73806 +	579.50	1086.68	3.5e-52	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AB22692 +	506.00	936.36	8.3e-44	4	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA69652 +	469.00	889.03	3.6e-41	53	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH15975 +	363.00	669.04	6.4e-29	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA160103 +	363.00	666.08	9.4e-29	27	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA321048 +	363.00	666.03	9.5e-29	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA202922 +	363.00	666.03	9.5e-29	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA754262 +	363.00	666.03	9.5e-29	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA91482 +	363.00	666.03	9.5e-29	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA294210 +	363.00	666.00	9.5e-29	27	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA158317 +	361.50	664.76	1.1e-28	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA18300 +	359.00	659.84	2.1e-28	23	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA01130 +	246.50	465.33	1.4e-17	30	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA74742 +	215.50	395.60	1.1e-13	7	
/SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:AA738368 +	114.50	187.79	0.0411	22	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA588411 +	111.00	179.85	0.1138	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA588533 +	111.00	179.85	0.1138	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AB207745 +	110.50	185.30	0.0566	1	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA244680 +	109.50	181.84	0.0882	16	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA74742 +	107.00	185.13	0.0579	7	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH05584 +	107.00	185.13	0.0579	7	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA105543 +	106.50	158.16	1.84	6	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA14503 +	105.00	158.12	1.85	6	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA18405 +	105.00	158.12	1.85	6	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA593080 +	104.00	165.87	0.6843	2	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AB22099 +	104.00	145.74	9.04	1	
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AB22098 +	104.00	143.59	11.91	1	

/SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:AA712324 + 102.50 159.77 1.50  
 /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA78084 + 101.00 162.22 1.09  
 /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:AA715847 + 100.50 161.16 1.25  
 /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA79056 + 100.00 156.18 2.37  
 /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AB22680 + 100.00 151.46 4.34  
 /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA746357 + 98.50 162.65 1.03

seq\_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA78001

seq\_documentation\_block:

ID AAX78001 standard; cDNA; 1669 BP.

AC AAX78001;

XX 19-AUG-1999 (first entry)

XX Potato GntI cDNA.

XX GntI; beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;  
 KW deficient; defective; detection; transgenic plant; sugar residue;  
 KW medicine; antigenicity; deglycosylation; potato; ds.

XX Solanum tuberosum.

XX Key Location/Qualifiers

FT CDS 53..1394

FT /\*tag= a

FT /product= "GntI"

XX DE19754622-A1.

PN 10-JUN-1999.

XX 09-DEC-1997; 97DE-1054622.

XX 09-DEC-1997; 97DE-1054622.

PA (VSC/ ) VON SCHAEWEN A.

PI Von Schaewen A;

XX WPI; 1999-338905/29.

DR P-PSDB; AAY08888.

XX Nucleic acid encoding plant N-acetylglucosaminyl transferase I  
 useful for generating plants producing glycosylated proteins of low  
 antigenicity

XX Claim 6; Page 16-18; 37pp; German.

XX This invention describes novel plant N-acetylglucosaminyl transferase I  
 (GntI) proteins and their encoding nucleic acids. The nucleic acid  
 sequences of the invention may be used for recombinant production of the  
 encoded proteins, which are then used to raise antibodies (Ab) for  
 identifying plants with deficient or defective production of the protein.  
 CC They may also be used to detect such plants by hybridization and to  
 CC isolate related sequences from other plants or to generate antisense or  
 CC sense constructs for reducing/deleting GntI protein activity in plants.  
 CC These transgenic plants may be used to produce glycoproteins with  
 CC minimal, uniform and defined sugar residues. Such glycoproteins are  
 CC useful in medicine and research, e.g. human glucocerebrosidase for  
 CC treating Gaucher's disease. Plants which are defective or deficient in  
 CC production of the GntI protein can be made to produce glycoproteins with  
 CC minimal, uniform and defined sugar residues, of low antigenicity. Use of  
 CC these plants eliminates the need for the difficult isolation and  
 CC deglycosylation of native proteins or preparation in defective animal  
 CC cells. This sequence represents potato (Solanum tuberosum) GntI encoding  
 CC cDNA.

XX Sequence 1669 BP; 489 A; 312 C; 387 G; 481 T; 0 other;

alignment\_scores:

Quality: 2345.00 Length: 446  
Ratio: 5.258 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-591-466C-2 x AAX78001 ..

Align seg 1/1 to: AAX78001 from: 1 to: 1669

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuLeuValValAl 17  
53 ATGAGAGGAACAAAGTTTGGCTTGTATTACGGTACCTTCTCGTGGC 102  
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34  
103 TGTCTCGCCCTTCATCATACATACAGATGCGGCTTTCGGCACACATCAG 152  
34 lutyryValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50  
153 AATATGTAGACCGCTTGTCTGTGCAATTGGAAGCAGAAAATCATTTGACA 202  
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlyArgVa 67  
203 AGTCAGACCATGTTGTTTGAACAAGATTAGCCAGCAGCAAGGAAGAGT 252  
67 lValAlaLeuGluGluGlnMetLysHisGlnAspGlnGlnCysArgGlnL 84  
253 AGTAGCTCTTGAAGCAAAATGAAGCATCAGGACCAAGGAGTGGCGCAAT 302  
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100  
303 TAAGGGCTCTTGTTCAGGATCTTGAAGTAAAGGCATATAAAAGTTAATC 352  
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117  
353 GGAGATGTGCAGATGCCAGTGGCAGCTGTAGTTGTTATGCGTTGCAGTCG 402  
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134  
403 TACTGTACTGTGAGAGGACTATTAAATCCATCTTAAATATACCAACAT 452  
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150  
453 CTGTTGCATCAAAATATCTCTTTTCATATCCAGGATGCATCAATCCT 502  
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167  
503 GATGTAAGAAAGCTTGCTTTGAGCTATGTTGTCAGCTGACGTATATGACGA 552  
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184  
553 CTGGATTATGAACCTGTGCATCTGNAAGACCAGGGAACTGGTTGCAT 602  
184 yrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200  
603 ACTACAAGATTGCACGTCAATACAAGTGGGCAITGGATCAGCTGTTTCAC 652  
201 LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAl 217  
653 AAGCATAAATTTTACCGGTGTTTCACTACTAGAAGATGATATGAAATTCG 702  
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234  
703 TGCTGATTTTTTTCATATTTTGGGCTGGAGCTACTCTTCTTGACAGAG 752  
234 spLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250  
753 ACAAGTCGATTATGGCTATTCTTCTTGGGAATGCAATGGACAAAGGCAG 802  
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe 267  
803 TTCGTCCAAGATCTGATGCTCTTTACCGCTCAGACTTTTTCCTGCTCT 852

267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpP 284  
853 TGGATGGATGCTTTCAAAATCAAACTTGGTCGAACTATCTCCAAAGTGC 902  
284 toLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300  
903 CAAGGCTTACTGGGATGACTGGCTAAGGCTCAAAGAAAAATCACAGAGT 952  
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHi 317  
953 CGACAAATTTATCGCCAGAGAGTTTGCAGAACGTACAATTTTGGTGAGCA 1002  
317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334  
1003 TGGTTCTAGTTTGGGGCAGTTTTTTAAGCAGTATCTTGAGCCAATTAAGC 1052  
334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350  
1053 TAAATGATGTCAGGTTGATTGGAAGTCAATGGACCTAAGTTACCTTTTG 1102  
351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367  
1103 GAGGACAACTATGTGAACACACTTTCGCGACTTGGTTAAAAAGGCTAAGCC 1152  
367 oIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384  
1153 CATCCACGGAGCTGATGCTGTTTTGAAAGCATTTAACATAGATGTTGATG 1202  
384 aIArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400  
1203 TCGGTATTTCAGTACAGAGACCAACTAGACTTTGAAGATATCGCTCGAC 1252  
401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrIly 417  
1253 TTTGGCATTTTGAAGAAATGGAAGGATGGTGACCACGGCGACATATAA 1302  
417 sGlyIleValValPheArgPheGlnThrSerArgArgValPheLeuValS 434  
1303 AGGGATAGTAGTTTTCCGGTTTCAACATCTAGACGTGTGTTCTTGT 1352  
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446  
1353 CCCCTGATTTCTCTGCACAACTTGGAGTTGAAGATACT 1390

seq\_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.AAX78002

seq\_documentation\_block:

ID AAX78002 standard; cDNA; 1737 BP.

XX AAX78002;

XX AC AAX78002;

XX DT 19-AUG-1999 (first entry)

XX DE Tobacco GntI cDNA.

XX KW GntI; beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;

XX KW deficient; defective; detection; transgenic plant; sugar residue;

XX KW medicine; antigenicity; deglycosylation; tobacco; ds.

XX OS Nicotiana tabacum.

XX FH Key Location/Qualifiers

XX FT CDS 127..1467

XX FT /\*tag= a

XX FT /product= "GntI"

XX PN DE19754622-A1.

XX PD 10-JUN-1999.

XX PF 09-DEC-1997; 97DE-1054622.

XX PR 09-DEC-1997; 97DE-1054622.



XX (VSCH/) VON SCHAEWEN A.  
 XX PI Von Schaewen A;  
 XX WPI: 1999-338905/29.  
 DR P-PSDB; AAY08889.  
 XX  
 PT Nucleic acid encoding plant N-acetylglucosaminyl transferase I  
 PT useful for generating plants producing glycosylated proteins of low  
 PT antigenicity  
 XX  
 XX Claim 7; Page 22-24; 37pp; German.  
 XX  
 CC This invention describes novel plant N-acetylglucosaminyl transferase I  
 CC (GntI) proteins and their encoding nucleic acids. The nucleic acid  
 CC sequences of the invention may be used for recombinant production of the  
 CC encoded proteins, which are then used to raise antibodies (Ab) for  
 CC identifying plants with deficient or defective production of the protein.  
 CC They may also be used to detect such plants by hybridization and to  
 CC isolate related sequences from other plants or to generate antisense or  
 CC sense constructs for reducing/deleting GntI protein activity in plants.  
 CC These transgenic plants may be used to produce glycoproteins with  
 CC minimal, uniform and defined sugar residues. Such glycoproteins are  
 CC useful in medicine and research, e.g. human glucocerebrosidase for  
 CC treating Gaucher's disease. Plants which are defective or deficient in  
 CC production of the GntI protein can be made to produce glycoproteins with  
 CC minimal, uniform and defined sugar residues, of low antigenicity. Use of  
 CC these plants eliminates the need for the difficult isolation and  
 CC deglycosylation of native proteins or preparation in defective animal  
 CC cells. This sequence represents tobacco (Nicotiana tabacum) GntI encoding  
 CC cDNA.  
 XX  
 SQ Sequence 1737 BP; 510 A; 337 C; 390 G; 500 T; 0 other;

alignment\_scores:  
 Quality: 2198.00 Length: 446  
 Ratio: 5.030 Gaps: 0  
 Percent Similarity: 97.982 Percent Identity: 92.825

alignment\_block:  
 US-09-591-466c-2 x AAX78002 ..

Align seg 1/1 to: AAX78002 from: 1 to: 1737

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17  
 127 ATGAGAGGGGAACAAAGTTTTCCTGTGATTTCCGGTACTCTCTCATCTGGC 176  
 17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34  
 177 TGTCTGTGCGCTTCATCTACACACAGATGCGGCTTTTTCGCACACAGTCAG 226  
 34 luTyrValAspArgLeuAlaAlaAlaIleGluAlaGluAsnHisCysThr 50  
 227 AATATGCAGATCGCTTGTGCTGCAATTGAACAGAAAATCAATTGTACA 276  
 51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67  
 277 AGCCAGACCATGCTTATTTGACCATGATTTAGCTGCAGCAGGAAGAAT 326  
 67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84  
 327 AGTTGCTCTTTGAAGACAAATGAAGCGTCAAGGACAGGAGTGCAGCAAT 376  
 84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100  
 377 TAAGGGCTCTTGTTCAGATCTTTGAAGATGAAGGCGCATAAAAAAGTTGATC 426  
 101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117  
 427 GGAATGTACAGATGCCAGTGGCTGTGTAGTTGTTATGCTTGAATCG 476

117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134  
 477 GGCTGATTACCTGGAAAGACATTAATAATCCATCTTAAATAACCAATAT 526  
 134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150  
 527 CTGTTGGCGTCAAAATATCTCTTTTCATATCCAGGATGGATCACAATCCT 576  
 151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167  
 577 GATGTCAGGAAGCTTGTCTTGAGCTATGATCAGCTGACGTATATGACGCA 626  
 167 steuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184  
 627 CTGGAATTTGAACCTGTGCATCTGAAAGACCGAGGAGCTGATTGCAT 676  
 184 yrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200  
 677 ACTACAAAATTCACGTCATTAACAAGTGGCATTTGGATCAGCTGTTTTAC 726  
 201 LysHisAsnPheSerArgValIleIleLeuGluAspMetGluIleAl 217  
 727 AAGCATTAATTTAGCGCTGTATCATACTAGAGATGATATGGAATTCG 776  
 217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234  
 777 CCTGATTTTGTGACTTTTGTGAGGCTGGAGCTACTCTCTTGACAGAG 826  
 234 sPlySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250  
 827 ACAAGTCGATTTATGGCTATTCTTCTTGAATGACAATGGACAATGCGAG 876  
 251 PheValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLe 267  
 877 TTGTCCAAAGATCCTTATGCTCTTTACCGCTCAGATTTTTTCCCGTCT 926  
 267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpP 284  
 927 TGGATGGATGCTTTCAAAATCTACTTTGGGACGAAATTTATCTCCAAGTGC 976  
 284 roLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300  
 977 CAAGGCTTACTGGGACGACTGGCTAAGACTCAAGAGAAATCACAGAGT 1026  
 301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHi 317  
 1027 CGACAATTTATTCGCCCAAGAAGTTTCAGAACATATAATTTTGGTCAGCA 1076  
 317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334  
 1077 TGGTCTTAGTTTGGGCGAGTTTTTCAAGCAGTATCTTGAGCCCAATTAAC 1126  
 334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350  
 1127 TAATGATGTCAGGTTGATTGGAAGTCAATGGACCTAGTTACCTTTTG 1176  
 351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367  
 1177 GAGGACAATTTACGTGAACACACTTTTGTGACTTGGTTAAAAAGGCTAAGCC 1226  
 367 oileHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384  
 1227 CATCCATGGAGCTGATGCTGCTTGAAGCATTTAAACATAGATGGTGATG 1276  
 384 aArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400  
 1277 TCGGTATTCAGTACAGATCACTAGACTTTGAAATATATCGCAGCGCAA 1326  
 401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaIaIaTyrLy 417  
 1327 TTTGGCATTTTGAAGAAATGGAAGGATGGTGTACCACGCTGCAGCATATAA 1376



Align seg 1/1 to: AAC39540 from: 1 to: 1639

```

9 AspLeuArgTyrLeuLeuValValAlaLeuAlaPheIleTyrIleGI 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 GACTTGAGATTCTCTCATCCGCGCAGCTTTCATGTTCATCATCATCA 99

25 nMetArgLeuPheAlaThrGlnSerGlnuYrValaspArgLeuAlaA 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 GATGAGGCTTTTCCAGCAGCAATCACAGATGCGATCCGCTCAGTTCG 149

42 laIleGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuLea 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 CTATCGAATCTGAGAACCATTCGACTAGTCAAATGCGAGCCTCATAG 199

59 LysIleSerGlnGlnGlyArgValValAlaLeuGluGluGlnMetly 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 GAAGTTAGCATCAACAGTCGCGGATTGTGCCCTCGAAGATGAAGAA 249

75 sHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValGlnAspLe 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 CCGCCAGGAGGAAGAACTTGTGCAGCTTAAGGATCTAATCCAGACG 299

92 luSerLysGlyIleLysLysLeuIleGlyAspValGlnMetProValAl 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 AAAAAAAGGAATAGCAAACTCACTCAAGGTGGCAGATGCTGTGGCT 349

109 AlaValValValMetAlaCysSerArgThrAspTyrLeuGluArgThr 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
350 GCTGTAGTGGTTATGCCCTGTCAGCTGTCGAGACTATCTTGAAGGAC 399

125 elysSerIleLeuLysTyrGlnThrSerValAlaSerLysTyrProLeu 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 TAAATCAGTTTAAACATATCAAACTCCGTTGCTTCAAAATATCCTAT 449

142 heIleSerGlnAspGlySerAsnProAspValArgLysLeuAlaLeuSe 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 TTATATCTCAGGATGGATCTGATCAAGCTGTCAAGAGCAAGTTCAG 499

159 TyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluProValHis 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
500 TATAATCAATTAAACATATGCAGCAGCTTGGATTTTGAACCAAGTGC 549

175 rGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyr 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
550 TGAAGGCTGCTGGAATGACTGCGTACTACAAGATGCGAGTGCCTACA 599

192 ystrpAlaLeuaspGlnLeuPheHisLysHisAsnPheSerArgValile 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
600 AGTGGGCACTGGACCAAGTTGTTTACAAACACAAATTTAGTCGAGT 649

209 IleLeuGluAspAspMetGluIleAlaAlaAspPheAspTyrPheGI 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
650 ATACTAGAAGATGATATGAAATTTGCTCCAGACTTCTTTGATTACTT 699

225 uAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSe 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 GGCTGCAGCTAGTCTCATGGATAGGATAAACCAATTATGGCTGCTAT 749

242 erTrpAsnAspAsnGlyGlnArgGlnPheValGlnAspProaspAlaLe 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
750 CATGGATGATATGACAGAGACAGCTTTGTGCATGATCCCTATCGCT 799

259 TyrArgSerAspPheProGlyLeuGlyTyrPmetLeuSerLysSerTh 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
800 TACCGATCAGATTTTTTCTGGCCTTGGGTGGATGCTCAAGAGATCGAC 849

275 rTrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspAspTrp 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
850 TTGGGATGAGTTATCACCAGAGTGGCCAAAGGCTTACTGGGATGATGG 899

292 euArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVal 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

alignment\_scores: Quality: 1776.00 Length: 438  
Ratio: 4.429 Gaps: 0  
Percent Similarity: 91.553 Percent Identity: 75.342

alignment\_block: ..  
US-09-591-466C-2 x AAC39540

```
900 TGAGACTAAAGGAAAAACCCATAAAGGCCGCCCAATTTCATTCGACCGGAAGTC 949
309 CysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePh 325
950 TGTAGAACATACAAATTTTGGTGAACATGGCTAGTTTGGGACAGTTTTT 999
325 eLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrpL 342
1000 CAGTCAGTATCTGGAACCTATAAAGCTAAACGATGTGACGGTTGACTGGA 1049
342 ySerMetAspLeuSerTyrLeuGluAspAsnTyrValLysHisPhe 358
1050 AAGCAAGACCTGGGATACCTGACAGAGGGAACCTATACCAAGTACTTT 1099
359 GlyAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLe 375
1100 TCTGGCTTAGTGAGACAAGCAGCAATTCAGGTTCTGACCTGTGCTT 1149
375 uLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAspGlnL 392
1150 AAAGCCTCAAAACATATAAGGATGATTCGTATCCGGTATAAAGACCAAG 1199
392 euAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLys 408
1200 TAGAGCTTTGACGCATTCGACGGGAATTTGGTATATTTGAAGAATGGAAG 1249
409 AspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPheG 425
1250 GATGCTGTCCTCGAACAGCATATAAAGGAGTAGTGGTTCGAATCCA 1299
425 nThrSerArgArgValPheLeuValSerProAspSerLeuArgGlnLeuG 442
1300 GACAAACAAGAGCTGTATTCCTGGTTGGGCCAGATTCTCTAATGACGCTG 1349
442 lValGluAspThr 446
1350 GAATTCGAATTC 1363

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC50197
seq_documentation_block:
ID AAC50197 standard; DNA; 1641 BP.
XX
AC AAC50197;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 63936.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
```



us-09-591-466c-2.rng

Wed Aug 14 11:48:26 2002

```
174 sThrGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisT 191
|||||
548 CACTGAAAGGCGCTGGCGAACTGACGCGTACTACAGATTGCACGCTCACT 597
|||||
191 YrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgVal 207
|||||
598 ACAAGTGGGCACTGGACCACTGTTTACAAACACAAATTTAGTCGAGTG 647
|||||
208 IleIleLeuGluAspAspMetGluIleAlaAlaAspPheAspTyrPh 224
|||||
648 ATTATACAGATGATATGGAATTTGCTCCAGACTTCTTTGATTACTT 697
|||||
224 eGluAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleS 241
|||||
698 TGAGCTGCACCTAGTCTCATGTAGGATAAAACCATATGCTGCTGCT 747
|||||
241 erSerTrpAsnAspAsnGlyGlnArgGlnPheValGlnAspProAspAla 257
|||||
748 CATCATGGAATGATAATGGACAGAAGCAGTTTGTGCATGATCCCTATGCG 797
|||||
258 LeuTyrArgSerAspPhePheProGlyLeuGlyTrpMetLeuSerLysS 274
|||||
798 CTATACCGATCAGATTTTTTCCCTGGCCTGGGTGGATGCTCAAGAGATC 847
|||||
274 rThrTrpSerGluLeuSerProLysTrpProLysAla.TyTrpAspAsp 290
|||||
848 GACTTTGGGATGAGTTATCACCAGAGTGGCCAAAGGCCCTTACTGGGATGAT 897
|||||
291 TrpLeuArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProG 307
|||||
898 TGGCTCAGACTAAGAGAAACCATAAAGCGCCCAATTCAATTCGACCGGA 947
|||||
307 uValCysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnP 324
|||||
948 AGTCTGTAACATACATTTTGGTGACATGGGTCTAGTTGGACAGT 997
|||||
324 hePheLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAsp 340
|||||
998 TTTTCAGTCAGTATCTGGAACCTATAAAGCTAAAGCATGTGACGGTTGAC 1047
|||||
341 TrpLysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHi 357
|||||
1048 TGGAAAGCAAGGACCTGGGTACCTGCAGACAGGGAACATAACCAAGTA 1097
|||||
357 sPheGlyAspLeuValLysAlaLysProIleHisGlyAlaAspAlav 374
|||||
1098 CTTTTCGCTTAGTGAGACAAGCAGCACAATTCAGGTTCTGACCTTG 1147
|||||
374 alLeuLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAsp 390
|||||
1148 TCTTAAAGGCTCAAAACATAAAGGATGATGTTGTTATCCGGTATAAAGAC 1197
|||||
391 GlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGlnTr 407
|||||
1198 CAAGTAGAGTTTGAAACCATTTGCAGGGGAATTTGGTATATTGGAAGAATG 1247
|||||
407 pLysAspGlyValProArgAlaIatYrLysGlyIleValValPheArgp 424
|||||
1248 GAAGGATGGTGTGCTCGAACACATATAAAGGAGTAGTGGTGTTCGAA 1297
|||||
424 heGlnThrSerArgArgValPheLeuValSerProAspSerLeuArgGln 440
|||||
1298 TCCAGACAACAAGAGCTGTATCTCTGTTGGGCCAGATTCTGTAATGCCAG 1347
|||||
441 LeuGlyValGluAspThr 446
|||||
1348 CTTGGAATTCGAATTC 1365
```

```
seq_name: /SID55/gcdata/geneseq/geneseq-emb1/NA1994.DAT:AAQ62625
seq_documentation_block:
ID AAQ62625 standard; cDNA to mRNA: 2557 BP.
XX
```

```
AC AAQ62625;
XX 18-JAN-1995 (first entry)
XX Rat N-acetyl-glucosamine transferase DNA.
XX production; detection: N-linked asparagine sugar chains;
XX Rat N-acetyl-glucosamine transferase; glycosylation; ds.
XX Rattus sp.
XX Location/Qualifiers
XX key 158..1501
XX CDS /tag= a
XX /product= N-acetyl-glucosamine transferase
XX JP06113841-A.
XX 26-APR-1994.
XX 08-OCT-1992; 92JP-0270299.
XX 08-OCT-1992; 92JP-0270299.
XX (ASAH ) ASahi Kasei Kogyo KK.
XX WPI; 1994-172739/21.
XX P-PSDB; AARS2657.
XX Rat N-acetyl-glucosamine transferase DNA and protein - useful in
XX prodn. and detection of enzyme
XX Claim 3; Page 10-12; 12pp; Japanese.
XX AAQ62625 encodes a rat N-acetyl-glucosamine transferase (see AARS2657).
XX The enzyme can be used when glycoproteins are produced in yeast so
XX that the natural sugar chain type can be added to the protein of
XX interest, ie. instead of the mannose residues that yeast normally
XX add.
XX Sequence 2557 BP; 535 A; 694 C; 718 G; 610 T; 0 other;
```

```
alignment_scores:
Quality: 735.50 Length: 422
Ratio: 2.599 Gaps: 13
Percent Similarity: 67.062 Percent Identity: 39.100
```

```
alignment_block:
US-09-591-466C-2 x AAQ62625 ..
```

Align seg 1/1 to: AAQ62625 from: 1 to: 2557

```
30 AlaThrGlnSerGluTyrValAspArgLeuAlaAlaIleGluAlaG 46
|||||
299 GCCAGCTCACCCCTGAGGTCATCCACCTGGCGAGGAGCCCGGCGGA 348
|||||
46 uAsnHisCysThrSerGlnThrArgLeuLeuIleAspLysIleSerGlnG 63
|||||
349 G.....TTGAGAGCGC 359
|||||
63 InGlnGlyArgValAlaLeuGluGlnMetLysHisGlnAspGln 79
|||||
360 AGCGGGGA.....CTACTGCAGCAGATCAAGGACCATATCT 397
|||||
80 GluCysArgGlnLeuArgAlaLeuValGlnAspLeuGluSerLysGlyI 96
|||||
398 CTGTGGAGGCGAG...AGCTGGAGAGTTCCACCCTGGCCCTCCAGCGTG 444
|||||
96 eLysLysLeuIleGlyAspValGlnMetProValAlaValValM 113
|||||
445 GCCCGGTGTGCTGGGACCCCTCACCAGCTGTGATCCCATTCGTGTCA 494
```

Length: 170

```
Ratio: 4.631          Gaps: 0
Percent Similarity: 92.353  Percent Identity: 80.588
alignment_block:
US-09-591-466c-2 x AA878003  ..
Align seg 1/1 to: AA878003  from: 1 to: 510

100 lileGlyaspValGlnMetProValAlaValValMetAlaCysSe 116
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
1  ATCGGAAGAGCTTGGATCCCGAGTGGCGCTGTAGTTGTATGGCTTCAG 50
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
116 rArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnT 133
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
51  TCGTGCAGACTACTCTGAAGAGGACTGTAAATCAGTTTAAACATATCAAA 100
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
133 hrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsn 149
|| : : : : : ||||||| : : : : : ||||||| : : : : :
101 CTCGCGTTCGTTCAAAATATCTCTATTTATATCTCAGGATGGATCTGAT 150
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
150 ProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetG1 166
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
151 CAAGCTGTCAGAGCAAGCTCATTGAGCTATATAATCAATTAACATATATGCA 200
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
166 nHisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuVala 183
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
201 GCACCTTGATTTTGAAGCAGTGTCTACTGAAGGCCTGGCGAAGTACTG 250
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
183 laTyrTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeuPhe 199
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
251 CGTACTACAGATGTCAGCTCACTACAGTGGGCACTGGACCACTGTTT 300
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
200 HisLysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluI1 216
: : : : : ||||||| : : : : : ||||||| : : : : :
301 TACAAGACAAATTTAGTCGAGTGATATATACTAGAAGATGATGGAAT 350
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
216 eAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspa 233
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
351 TGCTCCAGACTCTTGTATTACTTTGAGGCTGCAGCTAGTCTCATGGATA 400
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
233 rgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArg 249
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
401 GGGATAAAGCAATATATGGCTGCTTCATCATGAGTGCATGATGACAGAAG 450
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
250 GlnPheValGlnAspProAspAlaLeuTyrArgSerAspPheProG1 266
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
451 CAGTTTGTGCATGATCCCTATGCGCTATACCGCATCAGATTCTTCCCTGG 500
||||| : : : : : ||||||| : : : : : ||||||| : : : : :
266 yLeuGlyTrp 269
| : : : :
501 CCACGGCTGG 510
```

seq\_name: /STD55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:AAQ25594

seq\_documentation\_block:

ID AAQ25594 standard; cDNA; 2485 BP.

XX AAQ25594;

XX AC

XX DT

XX DE

XX 03-DEC-1992 (first entry)

XX Rabbit Gnt I cDNA clone.

XX High mannose glycoprotein; UDP-N-acetyl-glucosamine;

KW alpha-3-D-mannoside; beta-1,2-N-acetyl-glucosaminyl transferase I;

KW N-glycans; ss.

OS Oryctolagus cuniculus.

XX Key Location/Qualifiers

XX 50..1393

FT CDS

FT /tag= a

FT /product= Gnt\_I

```
FT polyA_signal 2435..2440
XX /*tag= b
PN W09209694-A.
XX
PD 11-JUN-1992.
XX
PF 29-NOV-1991; 91WO-CA00417.
XX
PR 30-NOV-1990; 90US-0620098.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Sarkar M, Schachter H;
XX
PI WPI; 1992-217073/26.
DR P-PSDB; AAR24779.
XX
XX Human and rabbit DNA sequences encoding GNT I enzymes - for
XX converting mannose to hybrid and complex N-glycan(s)
XX
XX Claim 3; Fig 4; 57pp; English.
XX
XX Rabbit liver Gnt I was purified and digested with pepsin, then
XX trypsin. The peptide fragments were sequenced and used to design
XX degenerate PCR primers (see AAQ30220-5). cDNA was prepared from total
XX RNA from rabbit liver. PCR was carried out on the cDNA preparation.
XX One of the two PCR products was cloned into the SmaI site of pGEM7z
XX for sequencing and then used as a riboprobe. The riboprobe was used
XX to screen a rabbit liver cDNA library in lambda gt10. The largest
XX insert in a positive clone was 1.6kb. An 80bp riboprobe was
XX prepared from the 5'-terminal of the 1.6kb insert and used to
XX rescreen the library. The largest cDNA insert was cloned into
XX pGEM-7z to obtain pGEM-7z-rcgntI. The full-length rabbit Gnt I
XX coding sequence was eventually obtained from overlapping clones.
XX See also AAQ25595.
XX
XX Sequence 2485 BP; 549 A; 623 C; 644 G; 669 T; 0 other;
```

alignment\_scores:

Quality: 726.50 Length: 400

Ratio: 2.721 Gaps: 13

Percent Similarity: 66.750 Percent Identity: 39.000

alignment\_block:

US-09-591-466c-2 x AAQ25594 ..

Align seg 1/1 to: AAQ25594 from: 1 to: 2485

82 ArgGlnLeuArgAlaLeuValGlnAspLeuGlu.....Se 93

||||| : : : : : ||||||| : : : : : ||||||| : : : : :

203 CBTGAGGTGATCGCTTAGCTCAGGATGCGAGGTAGTAGTTGGAACGTCA 252

93 rLysGlyIleLysLysLeuIleGlyAsp..... 102

||||| : : : : : ||||||| : : : : : ||||||| : : : : :

253 GCGGGGACTGTTGCACGACAGATTAGGAGCACCATGCTCTTTGGAGCCAGC 302

103 ..ValGlnMetProValAlaAla..... 109

||||| : : : : : ||||||| : : : : : ||||||| : : : : :

303 GGTGGAAGGTGCTACTGCGAGCCCTCTGCTCAGCCGCGATGTGCGCTGTG 352

110 .....ValValValMetAlaCysSerArgTh 118

||||| : : : : : ||||||| : : : : : ||||||| : : : : :

353 ACCCCACCCCGAGCTGTGTATCCCGCATCTCTGTAATTCCTGTGACCGCAG 402

118 rAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerV 135

||||| : : : : : ||||||| : : : : : ||||||| : : : : :

403 CACC...GTCCCGCGCTGTTTGGACAGAGCTACTGCATTATCGGCCCTCA. 448

135 a1AlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnProAsp 151

||||| : : : : : ||||||| : : : : : ||||||| : : : : :

449 ..GCTGAGCTGTTCCCGCATCATTTGTCAGCCAGCACTGTGGCATGAGGAG 496





02 ArgonLeu  
|||.....

1056 CGGAAGTGATTCCGCTGGCCCAAGACCGCGAGCTGGAGCTGGAGCGCAG 1105  
93 rLysGlyLeuLysLysLeuIleGlyAspVal.....G 104  
1106 CGTGGGCTGCTGCAGCAGATCGGGGATGCCCTGTCGAGCAGCGGGGA 1155  
104 InMetProValAlaAla..... 109  
1156 GGGTGCACCGCGCGCCCTCCCGCCACCGCGCTGTGCCTGTGACCCCC 1205  
110 .....ValValMetAlaCysSerArgThrAspTy 120  
1206 GCGCCGCGGTGATTCCTCCTGGTATCGCCTGTGACCGCAGCACT.. 1253  
120 rLeuGluArgThrLysSerIleLeuLysTyrGlnThrSerValAla 137  
1254 .GTTCGGCGCTGCTGCACAGCTGTGCATTATCGCCCTCTC...GCTG 1299  
137 erLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspValArg 153  
1300 AGCTCTTCCCATCATCTGTTAGCCAGCACTGCGGCACGAGGACGGCC 1349  
154 LysLeuAlaLeuSerTyrGlyGln....LeuThrTyrMetGlnHisLeuAs 169  
1350 CAGCCATCCCTCTACGGCAGCGGGTCCACCATCATCGGACGCGCGA 1399  
169 pTyrGlu.....ProValHisThrGluArgProGlyGluL 181  
1400 CCGTGCAGCAGCATTCGGTGCCCGCGCACCACTCCAGTTCACGGC.... 1445  
181 euValAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGln 197  
1446 .....TACTACAAGATCGCGCGCACTACCGCTGGCGCTGGCCAG 1487  
198 LeuPheHisLysHisAsnPheSerArgValIleIleLeuCluAspPhe 214  
1488 GTCTTCGGCAGCTTCGCTTCCCGCGCGGTGGTGGTGGAGGATGACCT 1537  
214 tGluIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuL 231  
1538 GGAGTGGCGCGGACTTCTCGAGTACTTCGGGCGCACTATCCGCTGC 1587  
231 euAspArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGly 247  
1588 TGAAGCCGACCCCTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1637  
248 GlnArgGlnPheVal.....GlnAspProAspAlaLeuTyrArgSerAs 262  
1638 AAGGACGAGATGGTGACGCCAGCAGCGCTGAGCTCTACCGCACCA 1687  
262 pPhePheProGlyLeuGlyTyrMetLeuSerLysSerThrTrpSerGluL 279  
1688 CTCTTTCCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1737  
279 euSerProLysTrpProLysAlaTyrTrpAspTrpLeuArgLeuLys 295  
1738 TGGAGCCCAAGTGGCAAGGCTTCTGGGACGACTGGATGGCGGCGCG 1787  
296 GluAsnHisArgGlyArgGlnPheIleArgProGluValCysArgThrTy 312  
1788 GAGCAGCGCAGGGCGGCTGTCATACGCTGAGATCTCAAGAACGAT 1837  
312 rAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnTyrL 329  
1838 GACCTTTGGCCCAAGGTGTGACCAACGCGGCGAGTCTTTGACCAACACC 1887  
329 euLysProIleLysLeuAsnAspValGlnValAspTrpLysSerMetAsp 345  
1888 TCAAGTTTATCAAGTGAACCAACGAGCTTGTGCATCTTCAACGAGTGCAC 1937  
346 LeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLeuVa 362  
1938 CTGTCTTACCTGCAGCGGGGCGCTATGACCGAGATTTC..... 1976

362 llLysLysAlaLysProIleHisGlyAlaAsp.....AlaValLeuLysA 377  
1977 .....CTGCGCGCGCTCTACGGTGTCCCGCAGCTCAGGTGGAGAAG 2019  
377 laPheAsnIleAsp.....GlyAspValArgIleGlnTyrArg 389  
2020 TGAGGACCAATGACCGAAGAGCTGGGGAGGTGGCGGTGCAGTATACG 2069  
390 AspGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluG 406  
2070 GCGAGGACAGCTTCAAGGCTTCGCCCAAGGCTCTGGGTGTATGGATGA 2119  
406 uTPpLysAspGlyValProArgAlaIaTyrLysGlyIleValValPheA 423  
2120 CCTTAGTGGGGTTCGAGAGCTGCTACCGGGGTATTGTCACTTC 2169  
423 rgPheGlnThrSerArgValPheLeuValSerPro 435  
2170 AGTTCGG...GCCCGCGGTGTCACCTGGCGCCCA 2204

seq\_name: /SID5/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.ABL22693

seq\_documentation\_block:

ID ABL22693 standard; DNA; 2367 BP.

XX ABL22693;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19552.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW Pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX PS WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Claim 1; SEQ ID NO 19552; 2lpp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (AB57737-AB577072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX CC Sequence 2367 BP; 601 A; 659 C; 581 G; 526 T; 0 other;

```

alignment_scores:
  Quality: 665.00      Length: 333
  Ratio: 2.737        Gaps: 10
  Percent Similarity: 72.973  Percent Identity: 43.544

alignment_block:
US-09-591-456C-2 x ABL22693  ..

Align seg 1/1 to: ABL22693 from: 1 to: 2367

106 ProValAlaValValValMetAlaCysSerArgThrAspTyrLeuG1 122
||||| |||||||||:|||||||:|||||||:|||||||:|||||||:
896 CCCGTGTTCCAGTCGTAGTCTTCGCCTCAATCGGGTGTGCG..GTGAA 942
122 uArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaSerLysT 139
:::~::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
943 GAAGTCAGTCGATAACTGGTTCAGTACAGGCCAGCCAGCTG...GAGCAGT 989
139 yrProLeuPheIleSerGlnAspGlySerAsnProAspValArgLysLeu 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
990 TCCCCATTATTGTTCCACAGGACTGGCGGATGACCCACCAGGAGGCA 1039
156 AlaLeuSerTyrGly...GlnLeuThrTyrMetGlnHisLeuAspTyrG1 171
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1040 ATCTCTCGTATGGCAACAGATCAGCTCATCGAGCAGCTGATCTCAG 1089
171 uProValHisThrGluArgProGlyGlu.....LeuValAlaTyrTyrL 186
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1090 CGACATC..ACAGTCTCCCGNAGGAGAAAAATTCAGAGCTACTACA 1138
186 ySleAlaArgHisTyrLysTyrPalaLeuAspGlnLeuPheHisLysHis 202
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1137 AGATACCCGACATCAGCGCTGGCGCCTGAACACCACTTC...GCTGTG 1183
203 AsnPheSerArgValIleIleLeuGluAspAspMetGluIleAlaAlaAs 219
:::||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1184 GGCTTTGAGTTCGTATATCTCGAGGACGATCTGAACGTGGCGCCGGA 1233
219 pPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLysS 236
||||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1234 CTTCCTTTGAGTACTTCCTGGGCAGCACAAAGTGTCTCAAGCAGGACCCCA 1283
236 erIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheVal 252
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1284 GCCTGTGTGCGTGTCGCGGTGGAATGCAATGCGCAGGCTGCTGTGTGTG 1333
253 Gln.....AspProAspAlaLeuTyrArgSerAspPheProGlyLe 267
:::||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1334 GACGCGCGCAGCCGAGCTGCTACCGCACCGATTTCTTTCCTCCGGTCT 1383
267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrp 284
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1384 CGCGTGGATGCTCACAAAGATCTGTGGCGGAGCTGCTCGGTCAAATGGC 1433
284 rOLysAlaIleTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisAtqGly 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1434 CCAAAATCTCTGGGATGATGGATACGTCATCCCGCCAGCCAGCAAGAT 1483
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluH1 317
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1484 CCGCTGTGCATTAGCCCGGAAATATCAGCAGCTCGCACGTTTTCGAAAAAT 1533
317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1534 AGCGCTTTCCAACGGTTTGTCTTCGATAAGTATCTGAAGCACATTAAAC 1583
334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
1584 TGACGGAGGACTTTCGTGAGTTTACAAAAATCAATATGACGTACCTGCTG 1633
351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1634 AAGGCAATTAACGATAACAGCTTCTTCGCGCGCGGTT...TATACGTATCC 1680

```

[illegible]

CC agents to down regulate expression and activity. The antibodies may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). Examples of diseases which may be treated include rheumatoid arthritis and diabetes.

XX  
SQ Sequence 2250 BP; 408 A; 661 C; 686 G; 495 T; 0 other;

alignment\_scores:  
Quality: 579.50 Length: 250  
Ratio: 3.167 Gaps: 5  
Percent Similarity: 73.200 Percent Identity: 46.000

alignment\_block:  
US-09-591-466C-2 x AAF93806 ..

Align seg 1/1 to: AAF93806 from: 1 to: 2250

194 AlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleLele 210  
400 AGCTCACCAGGCTCTCCGGCAGTTTCCTCCCGCGCGGTGGT 449  
210 uGluAspAspMetGluIleAlaAlaAspPhePheAspTyrPheGluAlag 227  
450 GGAGGATGACCTGGAGTGGCCCGGACTTCTTCGAGTACTTTCAGGCCA 499  
227 lYAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSerTrp 243  
500 CCTATCCGCTGTAAGCGCGACCCCTCCCTGCTGCTGCTCGGCCCTGG 549  
244 AsnAspAsnGlyGlnArgGlnPheVal.....GlnAspProAspAlaLe 258  
550 AATGACACGCAAGCAGCAGATGTTGGACGCCAGCCCTGACTGCT 599  
258 uTyrArgSerAspPheProGlyLeuGlyTrpMetLeuSerLysSert 275  
600 CTACCGCACCGACTTTTCCCTGGCTGGCTGGCTGGCTGGCTGGCTGG 649  
275 hrTrpSerGluLeuSerProLysTyrProLysAlaTyrTrpAspAspTrp 291  
650 TCTGGCTGAGTGGAGCCCAAGTGGCCCAAGGCTCTGGGACGACTGG 699  
292 LeuArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVa 308  
700 ATGCGCGCGCGGAGCAGCAGCGCGCGCGCGCTGCATACGCCCTGAGAT 749  
308 lCysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPheP 325  
750 CTCAGAAGCATGACCTTGGCCCGCAAGGGTGTGAGCCACGGCGAGTCT 799  
325 heLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrp 341  
800 TTGACGACGACCTCAAGTTATCAAGTGAACACGACGAGTTGTGCACTTC 849  
342 LysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisph 358  
850 ACCCAGCTGGACCTGCTTACCTGACCGGGGAGCCCTATGACCGAGATT 899  
358 eGlyAspLeuValLysLysAlaLysProIleHisGlyAlaAsp.....A 373  
900 C.....CTCGCCCGCTCTACGGTGTCTCCCGCTCCCGCTGCG 931  
373 laValLeuLysAlaPheAsnIleAsp.....GlyAspValArg 385  
932 AGGTGGACAAGTCAGGACCAATGACCGGAAGGAGTGGGGGAGTTCGG 981  
386 lIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGlnPheG 402  
982 GTGCAGTATACGGGACGAGCAGCTTCAAGGCTTTCGCCAAGGCTCTGG 1031  
402 YIlePheGluGlnTrpLysAspGlyValProArgAlaAlaTyrLysGlyI 419

1032 TGTCATGGATGACCTTAAGTCGGGGTTCGAGAGCTGGCTACCGGGTA 1081  
419 leValValPheArgPheGlnThrSerArgArgValPheLeuValSerPro 435  
1082 TTGTACCTTCCAGTTCGG...GGCGCGGTGTCCACCTGGCGGCCCA 1128

seq\_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:ABL22692

seq\_documentation\_block:  
ID ABL22692 standard; DNA; 4721 BP.  
XX  
AC ABL22692;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19549.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 19549; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4721 BP; 1175 A; 1113 C; 1154 G; 1279 T; 0 other;

alignment\_scores:  
Quality: 506.00 Length: 453  
Ratio: 2.057 Gaps: 20  
Percent Similarity: 54.305 Percent Identity: 32.450

alignment\_block:  
US-09-591-466C-2 x ABL22692/rev ..

Align seg 1/1 to reverse of: ABL22692 from: 1 to: 4721

106 ProValAlaAlaValValMetAlaCysSerArgThrAspTyrLeuGl 122  
2826 CCGGTGTTTCCAGTCGTAGTCTTCGCTGCAATCGGTGCTCG...GTGAA 2780

```
122 uArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaSerLysT 139
    :::: ||| ::::||||: ||| ||| ::::
2779 GAAGTCATCATGATTGTTTCAGTACAGGCCAGCGTG...GAGCACT 2733
    :|||:||||:||||: |||
139 yrProLeuPheIleSerGlnAspGly..... 147
    :|||:||||:||||: |||
2732 TCCCAATTATTGTGTACAGGATAGGATCGAAATGGGCTACCCAATCTAA 2683
    148 .....SerAsn..proAsp.....V 152
    ||||| |||
2682 ATCTTATACTTATCTATGCTCTAATACAGCACTCGCGCGATGAGCCCA 2633
    :|||:||||:||||: |||
152 aArgLysLeuAlaLeuSerTyrGly...GlnLeuThrTyrMetGlnHis 167
    :|||:||||:||||: |||
2632 CCAAGAGCAATCTCTCTGTTATGGCAACAGGTACAGCTCATCGAGCAG 2583
    168 LeuAspTyrGluProValHisThrGluArgProGlyGlu.....LeuVa 182
    ||| :::: ||| ||| |||
2582 CTGTACTCAGCGACATC...ACAGTGTGCCGAAGGAGAAAAATTCAA 2536
    182 lAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuP 199
    :|||:||||:||||: |||
2535 AGGCTACTACAAGATAGCCGACACTACGGCTGGGCCCTGAACACACCT 2486
    199 heHisLysHisAsnPheSerArgValIleIleLeuGluAspMetGlu 215
    || :::: ||| ||| ||| |||
2485 TC...GCTGTGGGCTTTGAGTTCTGTCATAATCTCGAGCAGCATCTGAAC 2439
    216 lIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAs 232
    :|||:||||:||||: |||
2438 GTGGCGCCGACTCTTTGAGTACTCTCTGGGCACGCACAACTGCTCAA 2389
    232 pArgAspLysSerIleMetAlaIleSerSerTrpAsnAspGlyGlnA 249
    :|||:||||:||||: |||
2388 GCAGGACCCCGCCTGCTGTGGTGTCCGCGTGAATGACAAATGCAAGG 2339
    249 rgGlnPheValGln.....AspProAspAlaLeuTyrArgSerAspPhe 263
    ||||| :|||:||||:||||: |||
2338 CTGCTGTTGTGACGCCCGCAGCGGAGTCTCTACCGCACCAGTTTC 2289
    264 PheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSe 280
    ||||| :|||:||||:||||: |||
2288 TTTCCCGGTCTCGGTGGATGCTACCAAGATCTGTGGCGCAGCTGTC 2239
    280 rProLysTrpProLys..... 285
    | ||||| |||
2238 GGTCAATGGCCCAATCGTGAGTGGGCAATATGTTGTTGCCAGCCA 2189
    286 .....AlaTyrTrpAspAspTrpLeuArg 293
    :|||:||||:||||: |||
2188 ATCGCTAATGGTTTCATTAACTGTTCTAGCTTCTGGGATGATTGGATAGT 2139
    294 LeuLysGluAlaAsnHisArgGlyArgGlnPheIleArgProGluValCysAr 310
    :|||:||||:||||: |||
2138 CATCCCGCCCGCAGCGCAAGATCGGTGTGCATTAGGCCCGGAAATATCAG 2089
    310 gThrTyrAsnPheGlyGluHisGlySerSer..Leu..... 321
    ||| :|||:||||:||||: |||
2088 CACTCGCAGTTTGAAAAAATAGCGGTTTCCAAAGTAAGAGATATAGCTAG 2039
    322 .....G 322
    :|||:||||:||||: |||
2038 CTAACAATGCCACGTCGCGCACTAATATGATAATTTGTTTAAACGCG 1989
    322 lGlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspValGln 338
    || ||||| :|||:||||:||||: |||
1988 GTTGTCTTCTCGATAAGTATCTGAAGCACATTAAACTGAGCGAGGACTTT 1939
    339 ValAspTrpLysSerMetAspLeuSerTyrLeuLeu..... 350
    |||:||||:||||: |||
1938 GTGCAGTTTACAAAAATCAATAGCTACCTGCTGAAGGTAGGTACACA 1889
    351 .....GluAspA 353
```

```
1888 CCCGTGCATTTGCTCATTTACACTAACAACTATTTCATGTTTCAGGACA 1839
    :|||:||||:||||: |||
353 snTyr..ValLysHisPhe..... 358
    ||||| :::: |||
1838 ATTAGCATAAACACGTTTCTCGCGCGGTTTATACGTATCCCATTTGTTACG 1789
    359 .....GlyAspLeuValLysLysAlaL 366
    ||||| |||
1788 TACGATGAGCTCGGCGAAACCTGATTAGTGAGTTGATATATATATATAT 1739
    366 ysProIleHisGlyAlaAlaValLeuLysAlaPheAsnIleAspGly 382
    :|||:||||:||||: |||
1738 AGAAAGTA...TCTACATAATTCCAATTATGGTGTTCAGAAATCGAAGGT 1692
    383 AspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaAr 399
    ||||| :|||:||||:||||: |||
1691 CCAGTTCGCATTCAATACACTACTAGGGAGCAGTACAGCGGACAACATAA 1642
    399 gGlnPheGlyIlePheGluGluTrpLys..... 408
    :|||:||||:||||: |||
1641 GATGCTGGTCTTATGGATGATTTCAGGTAGATCAATCAATCTATGTGC 1592
    409 .....Asp..GlyValProArgAlaA 415
    ||||| |||
1591 TTGAAACTATGATTATTCTTAATCTTTGTAGAGCGGTGTTCCGCGACTG 1542
    415 lATyrLysGlyIleValValPheArgPheGlnThrSerArgArgValPhe 431
    ||||| :|||:||||:||||: |||
1541 CCTACCATGGCATGCTCCTTC...TACTACAAACAAAGCGCGGTGCAT 1495
    432 Leu 432
    |||
1494 CTG 1492
    :|||:||||:||||: |||
seq_name: /SDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC69652
seq_documentation_block:
ID AAC69652 standard; cDNA; 535 BP.
XX AAC69652;
XX
XX 30-JAN-2001 (first entry)
XX
XX Human acetylglucosaminyltransferase coding sequence.
XX
XX Cytostatic; vaccine; human; breast tumour; antigen; breast cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200052165-A2.
XX
XX 08-SEP-2000.
XX
XX 29-FEB-2000; 2000WO-US05431.
XX
XX 04-MAR-1999; 99US-0262505.
XX
XX 19-MAR-1999; 99US-0272886.
XX
XX 17-SEP-1999; 99US-0396313.
XX
XX (CORI-) CORIXA CORP.
XX
XX Lodes MJ;
XX
XX WPI: 2000-572184/53.
XX
XX P-PSDB; AAB28190.
XX
PT Breast tumor antigen polypeptides and polynucleotides, useful for
PT manufacturing vaccines and compositions for treating, diagnosing, and
PT monitoring breast cancer
XX
PS Claim 16; Fig 1; 140pp; English.
XX
```

CC The present invention relates to immunogenic portions of new human  
CC breast tumour antigens (AAB28183-28214) and their coding sequences  
CC (AAC69645-C69804). The breast tumour antigen polypeptides of the present  
CC invention and their coding sequences are useful for inhibiting the  
CC development of breast cancer in a patient. The breast tumour antigen  
CC polypeptides and polynucleotides may be used in vaccines and  
CC pharmaceutical compositions for treating breast cancer, and for  
CC diagnosing and monitoring the cancer. The present sequence is a coding  
CC sequence for the immunogenic portion for one such human breast cancer  
XX tumour antigen.  
SQ Sequence 535 BP; 87 A; 173 C; 183 G; 92 T; 0 other;

alignment\_scores:  
Quality: 469.00 Length: 174  
Ratio: 3.608 Gaps: 4  
Percent Similarity: 74.713 Percent Identity: 50.000  
alignment\_block:  
US-09-591-466C-2 x AAC69652 ..  
Align seg 1/1 to: AAC69652 from: 1 to: 535

158 SerTyrGlyGln...LeuThrTyrMetGlnHisLeuAspTyrGlu..... 171  
|||||||:  
24 TCCTACGGCAGCGGTGTCAGCAGATCCGCGAGCCGACCTGAGCAGCAT 73  
  
172 .....ProValHisThrGluArgProGlyGluLeuValAlaTyrT 185  
||| ||| : : : |||  
74 TGGCGTCCCGCGGACCAACCGCAAGTCCAGGCG.....TACT 111  
  
185 YrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeuPheHisLys 201  
|||||||:  
112 ACAAGATCGCGCGCCACTACCGTGGCGCTGGCGCAGGTCTTCGCGCAG 161  
  
202 HisAsnPheSerArgValIleLeuGluAspAspMetGluIleAla 218  
:  
162 TTTCGCTCCCGCGCGCGGTGGTGGAGATGACCTGGAGGTGCGCCC 211  
  
218 aAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspL 235  
GACTTCTTCGAGTACTTTCGGCGCCACTATCGCTGCTGAGGCGGAC 261  
  
235 ySeriLeuMetAlaIleSerSerTyrAsnAspAsnGlyGlnArgGlnPhe 251  
|||||||:  
262 CTTCCCTGTCGCTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 311  
  
252 Val.....GlnAspProAspAlaLeuTyrArgSerAspPheProG 266  
||| : : : |||: : : |||: : : |||: : : |||: : : |||: : : |||  
312 GTGAGCGCCAGCAGCGCTGAGCTGCTTACCGCAGCGACTTTTCCCTGG 361  
  
266 yLeuGlyTyrMetLeuSerLysSerThrTrpSerGluLeuSerProLys 283  
|||||||:  
362 CTTGGGCTGGCTCTTGTGGCGCAGCTGCTGGCTGAGCTGGAGCCCAAG 411  
  
283 rProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArg 299  
|||||||:  
412 GGCCAAAGGCTTCTGGGAGGACTGATGCGGCGCGCGGAGCAGCGGAG 461  
  
300 GlyArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGly 316  
|||||||:  
462 GGGCGGCGCTGCATACGCTGAGATCTCAAGAACGATGACCTTTGGCGG 511  
  
316 uHisGlySerSerLeuGlyGln 323  
: : : ||| |||  
512 CAAGGTGTGAGCCGCGGAG 533

seq\_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH15975  
seq\_documentation\_block:  
ID AAH15975 standard; cDNA; 2078 BP.  
XX

AC AAH15975;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:14588.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
XX  
PR 27-AUG-1999; 99JP-0300253.  
XX  
PR 11-JAN-2000; 2000JP-0118776.  
XX  
PR 02-MAY-2000; 2000JP-0183767.  
XX  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
WP1: 2001-318749/34.  
XX  
PR Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 14588; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
SQ Sequence 2078 BP; 467 A; 582 C; 569 G; 460 T; 0 other;

alignment\_scores:  
Quality: 363.00 Length: 361  
Ratio: 1.658 Gaps: 18  
Percent Similarity: 60.665 Percent Identity: 29.917  
alignment\_block:  
US-09-591-466C-2 x AAH15975 ..  
Align seg 1/1 to: AAH15975 from: 1 to: 2078  
103 ValGlnMetProValAlaAlaValValMetAlaCysSerArgThrAs 119

Wed Aug 14 11:48:26 2002

```

1248 CGAATGCGAAGATGAT.....GACTTCACACCTGGACCCAGCTTGC 1291
401 eGlyIlePheGluGluTriPlysAspGlyValProArgAlaAlaTyrLysG 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1292 CAAGTGCTCCATATCTGGACACCTGGATGTG...CGTGGCAACCATCGG 1338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
418 lylleValValPheArgPheGlnThrSerArgValPheLeuVal... 433
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1339 GCCTG.....TGGAGATTGTTTGGGAAGAAGAACCACTTCCTCGGTG 1382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
434 .....SerProAspSerLeuArgGln 440
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1383 GGGGTCCGGCTTCCCTCACTCACTCACTCACTCACTCACTCACTCACT 1415
seq_name: /SID55/gcgdata/geneseq/genesequ-emb1/NA2001A.DAT-AA160103
seq_documentation_block:
ID- AA160103 standard; CDNA: 2711 BP.
XX
AC AA160103;
XX
XX 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4092.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
XX WO200153312-A1.
PN
PD 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
PF
XX 21-JAN-2000; 2000US-0488725.
PR
XX 25-APR-2000; 2000US-0552317.
PR
XX 09-JUL-2000; 2000US-0598042.
PR
XX 19-JUL-2000; 2000US-0620312.
PR
XX 03-AUG-2000; 2000US-0653450.
PR
XX 14-SEP-2000; 2000US-0662191.
PR
XX 19-OCT-2000; 2000US-0693036.
PR
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR
XX P-PSDB; AAM40947.
DR
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PT
XX
XX Claim 1; SEQ ID NO 4092; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression.
CC

```



CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX

SQ Sequence 2711 bp; 608 A; 755 C; 777 G; 571 T; 0 other;

alignment\_scores:  
 Quality: 363.00 Length: 361  
 Ratio: 1.658 Gaps: 18  
 Percent Similarity: 60.565 Percent Identity: 29.917

alignment\_block:

US-09-591-466C-2 x AAI60103 ..

Align seg 1/1 to: AAI60103 from: 1 to: 2711

```

103 ValGlnMetProValAlaAlaValValMetAlaCysSerArgThrAs 119
1029 CTCATGTGCTGTGGCT.....GTCAATGCAGGGAACCGACCCAA 1069
119 pTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValA 136
1070 TTACTCTACAGGATGCTGGCTCTCTGCTTTCAGCCAGGGGTGTCTC 1119
136 laSerLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspVal 152
1120 CTCAGATGATAACAGTTTTCATT.....CACGGCTACTATGAGGAACCC 1163
153 ArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLeuAs 169
1164 ATGGATGTGGTGGCAGCTGTTTGGT.....CTGAGGGGCATCCA 1201
169 pTyrGluProValHisThrGluArgProGlyGluLeuValAlaTyrTyrL 186
1202 GCATCTCCCATCAGCATCAAGAAATGCC.....C 1230
186 ysIleAlaArgHisTyrLysTrpAlaLeuAsp.....GlnLeuPhe 199
1231 GCCTGTCTCAGCACTACAAGCCAGCCTCACTGCCACTTTCAACCTGTTT 1280
200 HisLysHisAsnPheSerArgValIleLeuLeuGluAspMetGluI 216
1281 CCGAGGCCAAGTTTGCT.....GTGGTTCTGGAAGAGGACCTGGACAT 1324
216 eAlaAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspA 233
1325 TGCTGTGGATTTTTCAGTTTCTGAGCCCAATCCACTCACTGGAGG 1374
233 rgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArg 249
1375 AGGATGACAGCCTGTACTGCATCTCTGCTGGAATGACCAGGGTATGAA 1424
250 GlnPheValGlnAspProAspAlaLeuTyrArgSerAspPheProG 266
1425 CACACGGCTGAGGCCACCACTACTGTACCGTGGGATGGAGACCATGCCTGG 1474
266 yLeuGlyTrpMetLeuSerLysSerThrTrp...SerGluLeuSerProL 282
1475 GCTGGGCTGGGTGCTCAGGAGTCTTGTACAAAGGAGGAGCTTGAGCCCA 1524
282 ysTrpPro.....LysAlaTyr...TrpAspAspTrpLeuArgLeu 294
1525 AGTGGCCTACACCGGAAAGCTCTGGGATTTGGACATGTGGATGGCGATG 1574
295 LysGlnAsnHisArgGlyArgGlnPheIleArgProGluValCysArgTh 311
1575 CCGTAACAACCGCGGGCGAGAGTGCATCCTCCTGACGTTTCCCGATC 1624
311 rTyrAsnPheGlyGluHisGlySerSerLeu...GlyGlnPhePheLysG 327

```

```

:||||:||||| ||| :||| :||| ||| ||| :||| :|||
1625 CTACCACCTTTGGCATCGTCGGCCTCAACATGAATGGCTACTTTTCACGAGG 1674
327 lnTyrLeuGluProIleLysLeuAsnAspValGln...ValAspTrpLys 342
||||:||||| ||| :||| :||| ||| ||| :||| :|||
1675 CCTACTTCAAGAACACAAAGTTCACACACGGTTCCAGGTGTCCAGCTCAGG 1724
343 SerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheG 359
:||||:||||| ||| :||| :||| ||| ||| :||| :|||
1725 AATGTGGAC.....AGTCTGAAGAAAGAGCTTATGAAGTGAAGTTCA 1768
359 yAspLeuValLysLysAlaLysProIle...HisGlyAlaAspAlaVal 375
||||:||||:||||: ||| :||| :||| ||| ||| :||| :|||
1769 CAGGCTGCTCAGTGAGGCTGAGGTTCTGGACCACACGAAGAACCCTTGTG 1818
375 euLysAlaPheAsnIleAspGlyAsp.....Val 384
:||||: ||| :||| :||| :||| :||| :||| :|||
1819 AAGACTCTTTCTCTCCACAGACAGAGGGCCACACCTACGTGGCCTTTATT 1868
385 ArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGlnPh 401
||||:||||| ||| :||| :||| ||| ||| :||| :|||
1869 CGAATGGAGAAAGATGAT.....GACTTCACCACCTGGACCCAGCTTGC 1912
401 eGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLysG 418
:||||: ||| :||| :||| :||| :||| :||| :|||
1913 CAAGTGCCTCCATATCTGGACCTGGATGTG...CGTGGCAACCATCGGG 1959
418 yIleValValPheArgPheGlnThrSerArgArgValPheLeuVal... 433
||||: ||| :||| :||| :||| :||| :||| :|||
1960 GCCTG.....TGGAGATTGTTTCGAAGAAGAACCCACTTCTCTGGTGGTG 2003
434 .....SerProAspSerLeuArgGln 440
||||| ||||| :||| :||| :||| :||| :||| :|||
2004 GGGGTCCCGGCTTCCCGCTACTCAGTGAAGAAG 2036

```

OM of: US-09-591-466C-2 to: GenEmbl:\* out\_format : pfs  
Date: Aug 13, 2002 10:11 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet.p2n.model -DEV=xlp  
-O=/cpn2.1/USPTO\_spool/US09591466/runat\_12082002\_095312.4260/app\_query.fasta\_1.511  
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-DELEX=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DLEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pts -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09591466.@CEN1\_1\_7593  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-591-466C-2  
Query length: 446  
Database: GenEmbl:\*  
Database sequences: 1797656  
Database length: 1873333701  
Search time (sec): 2163.780000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
Sequence					
gb_pl:STU249878	+ 2345.00	3793.95	5.8e-203	1641	! A249878 Solanum tuberosum mRNA
gb_pat:AS95053	+ 2345.00	3793.78	5.9e-203	1669	! AS95053 Sequence 1 from Patent
gb_pl:STU249879	+ 2219.00	3588.86	1.5e-191	1691	! A249879 Solanum tuberosum pat
gb_pl:NTV16832	+ 2201.00	3557.76	8.3e-190	2031	! Y16832 Nicotiana tabacum mRNA
gb_pl:NTA249883	+ 2198.00	3554.63	1.2e-189	1708	! A249883 Nicotiana tabacum mRN
gb_pat:AS95055	+ 2198.00	3554.63	1.3e-189	1737	! AS95055 Sequence 3 from Patent
gb_pl:NTA249882	+ 2197.00	3552.27	1.7e-188	1836	! A249882 Nicotiana tabacum mRN
gb_pl:STU249880	+ 1948.00	3149.63	4.5e-167	1429	! A249880 Solanum tuberosum mRN
gb_pl:ATH243198	+ 1781.00	2876.25	7.6e-152	1820	! A243198 Arabidopsis thaliana
gb_pl:ATH249881	+ 1768.00	2855.07	1.2e-150	1830	! A249881 Arabidopsis thaliana
gb_pat:AS95057	+ 1768.00	2854.94	1.2e-150	1854	! AS95057 Sequence 5 from Patent
gb_pl:ATP20D10	- 1133.50	1779.59	9.2e-91	147215	! AL035538 Arabidopsis thalian
gb_pl:ATCHR1V89	- 1133.50	1776.51	1.4e-90	199789	! AL161593 Arabidopsis thalian
gb_pl:ATFE22113	- 1131.50	1780.89	7.8e-91	93760	! AL035539 Arabidopsis thalian
gb_hcg:AF004082	- 886.00	1379.35	1.8e-68	120538	! AP004082 Oryza sativa chromo
gb_pat:E07296	+ 735.50	1173.65	5.2e-57	2546	! D16302 Rat mRNA for N-acetylgl
gb_ov:XLGNTI	+ 735.50	1172.99	5.2e-57	2557	! E07296 cDNA encoding rat liver
gb_to:AF087456	+ 732.50	1172.22	4.3e-57	2508	! Y16819 Xenopus laevis mRNA for
gb_to:MUSGLNACT	+ 732.00	1170.95	7.4e-57	1894	! AF087456 Mesocricetus auratus
gb_to:BC008629	+ 732.00	1167.56	1.1e-56	2651	! M73491 Mouse N-acetylglucosam
gb_hcg:AL0606829	+ 732.00	1167.39	1.2e-56	2695	! BC006629 Mus musculus, Similar
gb_ov:RABUDPAA	+ 732.00	1167.39	1.2e-56	2695	! L07037 Mus musculus (clone E31
gb_to:MUSTRANX	+ 726.50	1159.27	3.3e-56	2485	! AL606829 Mus musculus chromo
gb_ov:AB087457	+ 725.50	1163.84	1.8e-56	1344	! AF087457 Mesocricetus auratus
gb_to:CGU65792	+ 725.50	1163.84	1.8e-56	1344	! U65792 Cricetus griseus muta
gb_to:AF087457	+ 725.50	1163.84	1.8e-56	1344	! AK057194 Homo sapiens cDNA FL
gb_to:CGU65791	+ 723.50	1160.59	2.8e-56	3103	! AF343963 Cricetus griseus N-
gb_to:AF343963	+ 723.50	1160.59	2.8e-56	3103	! U65791 Cricetus griseus N-ac
gb_to:CGU65791	+ 723.50	1153.93	6.8e-56	2602	! M55621 Human N-acetylglucosam
gb_pr:HUMGLNAC	+ 723.50	1153.62	6.8e-56	2683	! BC003575 Homo sapiens, mannosi
gb_pr:BC003575	+ 723.50	1147.96	1.4e-55	4705	! M61829 Human alpha-1,3-mannosy
gb_pr:HUMUDPNA	+ 723.50	1147.96	1.4e-55	4705	! AC022413 Homo sapiens chromo
gb_hcg:AC023255	- 723.50	1110.30	1.8e-53	196840	! AC023255 Homo sapiens clone
gb_to:AF343964	+ 718.50	1152.46	7.9e-56	1344	! AF343964 Cricetus griseus mu
gb_to:AF343965	+ 715.00	1147.59	1.5e-55	1344	! AF343965 Cricetus griseus mu
gb_ov:XLGCLNAC	+ 701.00	1116.29	8.1e-54	2894	! X83975 X laevis mRNA for N-ac
gb_in:AF082010	+ 663.00	1054.76	2.2e-50	2828	! AF082010 Caenorhabditis elegan
gb_in:AF082010	+ 645.00	1030.53	4.9e-49	1719	! AF082010 Caenorhabditis elegan
gb_in:AF082012	+ 633.00	1013.67	4.2e-48	1322	! AF082012 Caenorhabditis elegan

gb\_pat:AX136203 + 579.50 921.36 5.9e-43 2250 ! AX136203 Sequence 125 from  
gb\_in:AF082011 + 550.00 874.69 2.3e-40 1982 ! AF082011 Caenorhabditis ele  
gb\_hcg:AC014285 - 506.00 767.87 2.1e-34 65730 ! AC014285 Drosophila melano  
gb\_in:AC009218 - 506.00 759.10 6.4e-34 156753 ! AC009218 Drosophila melano  
gb\_in:AC008347 - 506.00 758.56 6.9e-34 165519 ! AC008347 Drosophila melano  
seq\_name: gb\_pl:STU249878  
seq\_documentation\_block:  
LOCUS STU249878 1641 bp mRNA linear PLN 04-JAN-2002  
DEFINITION Solanum tuberosum mRNA for N-acetylglucosaminyltransferase I (GntI  
gene), clone Al.  
ACCESSION AJ249878  
VERSION AJ249878.1 GI:18076139  
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Solanum.  
REFERENCE 1 (sites)  
AUTHORS Wenderoth,I., Tjaden,J. and von Schaewen,A.  
TITLE Isolation and characterization of different plant  
N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional  
analyses in the Arabidopsis cgl mutant, and in potato and tobacco  
antitense plants  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1641)  
AUTHORS von Schaewen,A.  
TITLE Direct Submission  
JOURNAL Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie,  
Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,  
GERMANY  
FEATURES  
source  
Location/Qualifiers  
1..1641  
/organism="Solanum tuberosum"  
/variety="Desiree"  
/db\_xref="taxon:4113"  
/clone="Al"  
/tissue\_type="leaf mesophyll"  
/note="sink"  
39..1379  
/gene="GntI"  
39..1379  
/gene="GntI"  
/EC\_number="2.4.1.101"  
/function="initiates complex-type glycans in the Golgi"  
/codon\_start=1  
/evidence=experimental  
/product="N-acetylglucosaminyltransferase I"  
/protein\_id="CAC80697.1"  
/db\_xref="GI:18076140"  
/translation="MRGNKFCFLRLVLLVVAALAFIYIOMRLFATOSEYVDRLLAAATE  
AENHCVSQTRLLLDKISQOGRYVALLSEQMKHQDQCRLRALVQDLSESGIKKLIGD  
VOMPAAVLLMKSRTDYLERIKTSVASKYPLFISODGSPDVRKLSYGL  
QLTMYGHLDEYVHTERPGLVAYYKIAHYKALDQHPKHNFSSVILDEDMETAA  
DFDYFEAGATLLDRKSIAMISWINDGROFQVDPDALYRSDFFGLGMLSKSTW  
SELPKPKRAYNDWDLRLKENHGRGQFIRPEVCRTYNGEHSGSLGDFQYLEPIKL  
NDVQVDMKSNLISYLLLENNYKFGOLVKKAKPIHGADAVLKAFLNIDGVDVYQYRQDL  
DPEDIAHQGFEEYMRKDGVPYRAIKGVVFRFQTSRRVFLVSPDSLRQLGVEDT"

BASE COUNT 485 a 302 c 377 g 477 t  
ORIGIN  
alignment\_scores:  
Quality: 2345.00 Length: 446  
Ratio: 5.258 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-591-466C-2 x STU249878 ..  
Align seg 1/1 to: STU249878 from: 1 to: 1641

```
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuLeuValValAl 17
|||||
39 ATGAGAGGGAACAAGTTTTCGTTGATTACGGTACCTTCGTCGTGGC 88
|||||
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
89 TGGCTCTGGCTTCATCATCATACAGATGCGGCTTTTCGCGACACAGTCAG 138
|||||
34 LuTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
139 AATATGTAGACCCCTTGCTGTCGAATTCAGCAGAAAAATCATTTGTACA 188
|||||
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
|||||
189 AGTCAGACCAAGATTGCTTATTCACAAGATTAGCCAGCAGCAAGGAAGGT 238
|||||
67 lValAlaLeuGluGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
239 AGTAGCTCTTTGAAGAACAAATGAAGCATCAGGACCAGGAGTGCCGCAAT 288
|||||
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
289 TAAGGCTCTGTTTCAGGATCTTTGAAGTAAGGCATAAAAAGTTAATC 338
|||||
101 GlyAspValGlnMetProValAlaAlaValValValMetAlaCysSerAr 117
|||||
339 GGAGATGTGCAGATGCCAGTGGCAGCTGTAGTTGTTATGGCTTGCAGTCG 388
|||||
117 qThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
389 TACTGACTACTCGAGAGGACTATTAAATCCATCTTAAATAACCAACAT 438
|||||
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
439 CTGTGTCATCAAAATATCTCTTTTCATATCCAGGATGGATCAAAATCCT 488
|||||
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
489 GATGTAAGAAAGCTTGCTTTGAGCTATGTCAGCTGACGTATATGCAGCA 538
|||||
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
|||||
539 CTTGGATTATGAACCTGTGCATACTGAAAGACCAGGGAACCTGTTGCAT 588
|||||
184 yTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeuPheHis 200
|||||
589 ACTACAAGATTGCCAGTCATTAAGTGGGCAATGGATCAGCTGTTTCAC 638
|||||
201 LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAl 217
|||||
639 AAGCAATAATTTAGCCGTGTTATCATACTAGAAGATCATATGAAATTCG 688
|||||
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArg 234
|||||
689 TGGTGAATTTTGGACTATTTTGGAGCTGGAGCTACTCTTCTTGCAGAG 738
|||||
234 sPlySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
|||||
739 ACAAGTCGATTATGGCTATTTCTTCTTGAATGCAATGCACAAAGGCAG 788
|||||
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe 267
|||||
789 TTGCTCCAAGATCCTGATGCTCTTTACCGCTCAGACTTTTTCCTGCTCT 838
|||||
267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpp 284
|||||
839 TGGATGATGCTCTTCAAAATCAACTTGGTCCGCAACTATCTTCCAAAGTGC 888
|||||
284 roLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300
|||||
889 CAAAGGCTTACTGGGTGACTGCTGAAGCTGAAGAAATAATCACAGAGGT 938
|||||
```

```
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHi 317
|||||
939 CGACAATTTATTCGCCCAGAAAGTTTCGAGAAGCTACAAATTTTGTGTGACA 988
|||||
317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334
|||||
989 TGGTCTTAGTTTGGGCAGTTTTTTAAGCAGTATCTTCAGCCCAATTAAGC 1038
|||||
334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350
|||||
1039 TAAATGATGTCAGGTTGATTGGAAGTCAATGGACCTAAGTTACCTTTTG 1088
|||||
351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367
|||||
1089 GAGGACAACTATGTGAACACTTTGGCGACTTGGTTAAAAAGGCTAAGCC 1138
|||||
367 oIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384
|||||
1139 CATCCAGGAGCTGATGCTCTTTTGAAGCATTTAATAGATGTCATG 1188
|||||
384 aArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400
|||||
1189 TGGGTATTTCAGTACAGAGCAACTAGACTTTGAAGATATCGCTCGACAG 1238
|||||
401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLy 417
|||||
1239 TTTGGCATTTTGAAGAATGAAGGATGGTGTACCCAGCGGCAGCATATA 1288
|||||
417 sGlyIleValValPheArgPheGlnThrSerArgArgValPheLeuValS 434
|||||
1289 AGGATAGTAGTTTTCGGTTTTCAACATCTAGACGTGTGTTCTCTGTTT 1338
|||||
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
|||||
1339 CCCTGATCTCTTCGACAACCTGGAGTTGAAGATACT 1376
|||||
seq_name: gb_pat:A95053
seq_documentation_block:
LOCUS A95053 1669 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 1 from Patent WO9929879.
ACCESSION A95053
VERSION A95053.1 GI:6779205
KEYWORDS .
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 1669)
AUTHORS Von,S.A.
VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH
A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GNTI)
ACTIVITY
Patent: WO 9929879-A 1 17-JUN-1999;
VON SCHAEWEN ANTJE (DE)
FEATURES
Location/Qualifiers
source
1..1669
/organism="Solanum tuberosum"
/strain="DESIREE"
/db_xref="taxon:4113"
/cell_type="BLATTZELLEN"
/tissue_type="MESOPHYLL"
/clone_lib="LAMBDA ZAP II (ECORI)"
/dev_stage="(B) CLON(E): GNTI-AL(K)"
15..52
53..1393
/gene="CGL"
53..1393
/gene="CGL"
/standard_name="GNTI"
/EC_number="2.4.1.101"
```

```

/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.
GLYKOPROTEINEN"
/notes="ERSTE GNTI-SEQUENZ AUS KARTOFFEL (UNPUBLIZIERT)"
/codon_start=1
/evidence=experimental
/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE 1"
/protein_id="CAB70462.1"
/db_xref="GI:6779206"
/translation="MRGNFCFDRLRYLLVVAALAFIYIOMRLFATQSEYVVDRLAAAE
AENHFSQIRLLIDKISQOQGRVVALEQMKHQDQCRQLRALVDLESKGIKKLIGD
VQMPAAVVMACSRDYLERITKSILKYQTSVASKYPLIFISQDGSNPDVRKLAISYG
QNTYMOHLDEVPVHTERPGELVAYYKIAIRHYKWLDPDLFKHNFSRVILIEDDMETAA
DFDYPEAGATLLDRDKSITMAISSNDNGOROFVODPDALYRSDFEFGLGWMLSKSTW
SELSPKWPKAYDDMLRLKENHGRQFIRPEVCRTYNFEHGSGLSQFFKQXLEPKL
NDVQDPKSMDSLLELYEDNYKHFGLVKKARPIHGADAVLAFNIDGVRLOYRDQL
FEDIAQRGIFEEWKGVPRAAYKGIWVRFQTSRRVFLVSPDSLRLQIGVED"
<80. .>139
CDS
/genes="CGL"
/standard_name="MEMBRANANKER EINES TYP II GOLGI-PROTEINS"
/notes="MEMBRANANKER (AMINO ACID 10-29)"
/function="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN
IDENTIFIZIERT"
/codon_start=1
/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"
/protein_id="CAB70463.1"
/db_xref="GI:6779207"
/translation="LRYLLVVAALAFIYIOMRLF"
1394. .1655
489 a 312 C 387 g 481 t
3'UTR
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 2345.00      Length: 446
  Ratio: 5.258          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-591-466C-2 x A95053  ..

Align seg 1/1 to: A95053 from: 1 to: 1669

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17
|||||
53 ATGAGAGGGACAGATTTCCTTGATTACGGTACCTCTCGTCGGGCG 102

17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
103 TGCTCTCGCCTTCATCTACATACAGATGCGGCTTTTCGGCACACAGTCAG 152

34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
153 AATATGTAGACCGCCTTGCTGCTGCAATTGAAGCAGAAAAATCATTTGACA 202

51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
|||||
203 AGTCAGACACAGATTGCTATTATGACAGATTAGCCACGACGAGAGAGT 252

67 lValAlaLeuGluGluClnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
253 AGTAGCTCTTTGAAGAACAAATGAAGCATCAGGACACAGAGAGTGCAGCAAT 302

84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
303 TAAGGGCTCTTGTTACAGGATCTTGAAGTTAAGGCCATATAAAGTTAATC 352

101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
|||||
353 GGAGATGTGCAGATGCCAGTGGCAGCTGTAGTTGTTATGGCTTGCAGTCG 402

117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
403 TACTGACTACCTGGAGAGGACTATTAAATCCATCTTAAATAATACCAACAT 452

```

```

134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
453 CTGTTTGCATCAAAATATCTCTTTTCATATCCAGAGTGGATCAAAATCCT 502

151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHl 167
|||||
503 GATGTAAAGAAAGCTTGGCTTTTGAGCTATGGTCAGCTGATATATGCAGCA 552

167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
|||||
553 CTTCGATTATGAACCTGTGCATCTAGAAAGACCAAGGGAAGTGGTTGCAT 602

184 yTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
|||||
603 ACTACAAGATTGCAAGCTCATACAAATGGGCATTGGATTCAGCTCTTTTCC 652

201 LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAl 217
|||||
653 AAGCAATATTTTACCCGTGTATCATCTAGAAATGATATGGAAATTGC- 702

217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234
|||||
703 TGCTGATTTTTTGGACTATTTTGGAGGCTGGAGCTACTCTTCTTGACAGAG 752

234 sPlysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
|||||
753 ACAAGTCGATTATGGCTATTCTTCTTGGAAATGACAATGGACAAGGCAG 802

251 PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe 267
|||||
803 TTGCTCCAAGATCCTGATGCTCTTTACCGCTCAGACTTTTTTCTCGTGCT 852

267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrp 284
|||||
853 TGGATGGATGCTTTCAAAATCAACTTGGTCCGAACATATCTCCAAGTGC 902

284 roLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300
|||||
903 CAAGGCTTACTGGGATGACTGGCTAAGGCTGAAGAAAAATCACAGAGT 952

301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHl 317
|||||
953 CGACAATTTATTCGCCAGAAAGTTTCAGAAACGTACAATTTTGGTGAGCA 1002

317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334
|||||
1003 TGTTCTTAGTTTGGGGCAGTTTTTTAAGCAGTATCTTGAGCCAAATTAAGC 1052

334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350
|||||
1053 TAAATGATGTCCAGGTTGATTGGAAGTCAATGGACCTAAGTTACCTTTTG 1102

351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367
|||||
1103 GAGGACAACATATGTGAACACACTTTGGCGACTTGGTTTAAAAAGGCTAAGCC 1152

367 otleHisGlyAlaAspAlaValIleLysAlaPheAsnIleAspGlyAspV 384
|||||
1153 CATCCAGGAGCTGATGCTGTTTGAAGCATTTAACATAGATGGTGATG 1202

384 aArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400
|||||
1203 TCGGTATTCAGTACAGACACCAACTAGACTTTGAAGATATCGCTCGACAG 1252

401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLy 417
|||||
1253 TTTGGCATTTTGAAGATGGAAGGATGGTGACCACGGCAGCATATAA 1302

417 sGlyIleValValPheArgPheGlnThrSerArgValPheLeuValS 434
|||||
1303 AGGGATAGTAGTTTCCCGGTTTCAAAACATCTAGACGTGTGTTCTTCTT 1352

```

434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446  
 1353 CCCCTGATCTCTTCGCAACTTGGAGTTGAAGATACT 1390  
 seq\_name: gb\_pl:STU249879

seq\_documentation\_block:  
 LOCUS STU249879 1691 bp mRNA linear PLN 04-JAN-2002  
 DEFINITION Solanum tuberosum partial mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A6.

ACCESSION AJ249879  
 VERSION AJ249879.1 GI:18076141  
 KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.  
 SOURCE potato.

ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (sites)  
 AUTHORS Wenderoth, I., Tjaden, J. and von Schaewen, A.  
 TITLE Isolation and characterization of different plant  
 N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional  
 analyses in the Arabidopsis cgl mutant, and in potato and tobacco  
 antisense plants  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1691)  
 AUTHORS von Schaewen, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie,  
 Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,  
 GERMANY

#### FEATURES

Source  
 1..1691  
 /organism="Solanum tuberosum"  
 /variety="Desiree"  
 /db\_xref="taxon:4113"  
 /clone="A6"  
 /tissue\_type="tuber"  
 /note="sink"  
 1..1298  
 /gene="GntI"  
 1..1298  
 /gene="GntI"  
 /EC\_number="2.4.1.101"  
 /function="initiates complex-type glycans in the Golgi"  
 /codon\_start=3  
 /product="N-acetylglucosaminyltransferase I"  
 /protein\_id="CAC80698.1"  
 /db\_xref="GI:18076142"  
 /translation="VAALAFIYIQMRIFATOSEYVDRIAAAEADHCTSQTRLLIDK  
 ISQOGRVVALEEMKQDQECRLALVODLESKGIKKLIGDVQMPVAAVVMACSR  
 ADYLEKTKSLKYQTSVASKYPLFISQDGNPDVRKLALSYDQLTYMHLDFFPVHT  
 ERPELVAYYKIAHHYKWDLPFKHNFSRVILEDDMETAADFFDYFEAGATLLDR  
 DKSIMAISSNDNGORQFVQDPYALYRSDFPGLGWLMSKSTWSELSPKPKAYWDDW  
 LRLKNHRRGROFIRPEVCRTYNFGHGSSLGQFFKQVLEPIKLNQVDVQWKSMDLSYL  
 LEDNVYHFGDLVAKKPIHGADAVLAFNFDGVRVQYRQDLDFDIARQFIFEEM  
 KDGVPRAAYKGIYVFRQTSRRVFLVAPDSLRQLGVEDT"  
 BASE COUNT 518 a 307 c 381 g 485 t  
 ORIGIN

alignment\_scores:  
 Quality: 2219.00 Length: 431  
 Ratio: 5.172 Gaps: 0  
 Percent Similarity: 99.536 Percent Identity: 98.144

#### alignment\_block:

US-09-591-466C-2 x STU249879 ..  
 Align seg 1/1 to: STU249879 from: 1 to: 1691  
 16 valAlaAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrG1 32  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

3 GTGGCTGCTCTCGCCTTCATCTACATACAGATCGCGCTTTTCGGCAGACACA 52  
 32 nSerGluTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisC 49  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 53 GTCAGAAATATGTAGACCCCTTGCTGCTGCAATTTGAAGCAGCAAGATCAT 102  
 49 ySthrSerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGly 65  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 103 GTACAAGTCAGACAGATTCCTTATTGACAAGATTATGCCAGCAGCAAGGA 152  
 66 ArgValValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysAr 82  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 153 AGAGTAGTGGCTCTTGAAGAACAAATGAAGCGTCAGGACGAGAGTCCG 202  
 82 gGlnLeuArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysL 99  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 203 GCAATTAAGGGCTCTTGTTCAGGATCTTTGAAGTAAGGCGCATAAAAAGT 252  
 99 euIleGlyAspValGlnMetProValAlaAlaValValMetAlaCys 115  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 253 TAATCGGAGATGTCAGATGCCAGTGGCGCTGTAGTTGTTATGGCTTGC 302  
 116 SerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrG 132  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 303 AGTCGTGCTGACTACCTGGGAGAGACTATTAAATCCATCTTAAATACCA 352  
 132 nThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerA 149  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 353 AACATCTGTTGCATCAAAATATCCTCTTTTCATATCCAGATGGATCAA 402  
 149 snProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMet 165  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 403 ATCCTGATGTTAGAAAGCTTGGCTTTGAGCTATGATCAGCTGACGTATATG 452  
 166 GlnHisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuVa 182  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 453 CAGCACTTGGATTTTGAACCTGTGCATCTGAAAGACCGGGAAGTGGT 502  
 182 lAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuP 199  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 503 TGCATACTACAAGATTGCAGCTATTACAAGTGGGCACTTGGATCAGCTGT 552  
 199 heHisLysHisAsnPheSerArgValIleIleLeuGluAspAspMetGlu 215  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 553 TTCACAAGCATATATTTAGCCGTGTTATCATCTACTAGAAGATGATATGGA 602  
 216 lIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAs 232  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 603 ATTGCTGCTGATTTTGTGACTATTTTGAGGCTGGAGCTACTCTTCTTGA 652  
 232 pArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnA 249  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 653 CAGAGATAAGTCGATTATGGCTATTTTCTCTTGGAAATGACAATGGACAAA 702  
 249 rgGlnPheValGlnAspProAspAlaLeuTyrArgSerAspPhePro 265  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 703 GGCAGTTCGTCCTCAAGATCCTTATGCTCTTTACGCTCAGACTTTTTCCT 752  
 266 GlyLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLy 282  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 753 GGTCTTGGATGATGCTTTCAAAATCAACTTGTGCGAACTATCTCCAAA 802  
 282 sTrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisA 299  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 803 GTGGCCAAAGGCTTACTGGGATGACTGGCTAAGGCTCAAGGAAATACACA 852  
 299 rgGlyArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGly 315  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 853 GAGGTGCAAAATTTATTCGCCCCAGAGTTTGCAGAACATACAAATTTTGT 902  
 316 GluHisGlySerSerLeuGlnPhePheLysGlnTyrLeuGluProI1 332  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 903 GAGCATGTTCTAGTTTGGGCGAGTTTTTTTAAAGCAGTATCTTGAGCCAAT 952

332 eLysLeuAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrL 349  
|||||  
953 TAAGCTAAATGATGTCAGGTTGATTGGAAGTCAATGGACCTAAGTTACC 1002  
349 euLeuGluAspAsnTyrValLysHisPheGlyAspLeuValLysAla 365  
|||||  
1003 TTTTGGAGGACAACATATGTGAACACATTTGGCGACTTGGTTAAAGGCT 1052  
366 LysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspG 382  
|||||  
1053 AAGCCCATCCACGGTGTGATGCTGTTTGAAGACATTTAAACATAGATGG 1102  
382 YAspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAla 399  
|||||  
1103 TGATGCGGTATTTCAGTACAGACCACTAGACTTTGAAGATATCGCTC 1152  
399 rgGlnPheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAla 415  
|||||  
1153 GGCAGTTTGGCATTTTGAAGAATGAAGATGCTGTACCACGGCAGCA 1202  
416 TyrLysGlyIleValValPheArgPheGlnThrSerArgValPheLe 432  
|||||  
1203 TATAAAGGGATAGTACTTTCGGGTTTCAACATCTAGACGTGTGTTCT 1252  
432 uValSerProAspSerLeuArgGlnLeuGlyValGluAspThr 446  
|||||  
1253 TGTTGCCCTGATTCTCTCGACAACCTTGAGTTGAAGATACG 1295

seq\_name: gb\_pl:NTY16832

seq\_documentation\_block:  
LOCUS NTY16832 2031 bp mRNA linear PLN 24-AUG-1999  
DEFINITION Nicotiana tabacum mRNA for beta-1,2-N-acetylglucosaminyltransferase, complete CDS.  
ACCESSION Y16832  
VERSION Y16832.1 GI:5764048  
KEYWORDS beta-1,2-N-acetylglucosaminyltransferase.  
SOURCE common tobacco.  
ORGANISM Nicotiana tabacum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
1 (bases 1 to 2031)  
Strasser,R., Mucha,J., Schwiela,H., Altmann,F., Glossl,J. and  
Steinkellner,H.  
Molecular cloning and characterization of cDNA coding for beta1,  
2N-acetylglucosaminyltransferase I (GlcNAc-TI) from nicotiana  
tabacum  
Glycobiology 9 (8), 779-785 (1999)  
99335389  
2 (bases 1 to 2031)  
Steinkellner,H.  
Direct Submission  
Submitted (06-MAR-1998) H. Steinkellner, Zentrum fuer Angewandte  
Genetik, BOKU-Wien, Muthgasse 18, 1190 Wien, AUSTRIA  
Location/Qualifiers  
1. .2031  
/organism="Nicotiana tabacum"  
/cultivar="xanthi"  
/db\_xref="taxon:4097"  
1. .76  
77. .1417  
/EC\_number="2.4.1.101"  
/codon\_start=1  
/product="alpha-1,3-mannosyl-glycoprotein  
beta-1,2-N-acetylglucosaminyltransferase"  
/protein\_id="CAB53347.1"  
/db\_xref="GI:5764048"  
/translation="MRGYKPCDFRYLLILAFAVLYIOMRLFATOSEYADRLAAALE  
AENHCTSTQRLLDQISQOGRIVALEQMKRQDCRQLRALVQDLESKGRKLIGN  
VQMPVAAYVMACNRADYLEKTSIKLYQISVAPKYPPIFISQDGHDPVRRKLALSYD  
QLTYMQLDFEPVHTERPGEIJAYYKIARHYKWLQDLFYKHNFSRVIILEDDEIAP

5'UTR  
CDS

DDFFFEAGATLLDRDKSIMAIISSHNDCQMDFVQDPYALYRSDFPGLGWLMSKSTW  
DELSPWPKAYMDDWLRLKENHGRQFIRPEVCRSTNFGEHSSLGQFFKQYLEPIKL  
NDVQVDMKSMDSLYLLEDNLYVKHAKPIHGADAVLAKAFNIDGVRLOYRDOL  
DFEDTARQFGIFEENKDGVPRAAYKGIYVFRYQTSRRVLVGPDSLIQULQUNEDT\*  
1418. .2031  
3'UTR 606 a 380 c 427 g 618 t  
BASE COUNT  
ORIGIN  
alignment\_scores:  
Quality: 2201.00 Length: 446  
Ratio: 5.037 Gaps: 0  
Percent Similarity: 97.982 Percent Identity: 93.049  
alignment\_block:  
US-09-591-466c-2 x NTY16832 ..  
Align seg 1/1 to: NTY16832 from: 1 to: 2031  
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17  
|||||  
77 ATCAGAGGGGTACAAGTTTTCGTGTGATTTCGGGTACCTCTCATCTTGGC 126  
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34  
|||||  
127 TCGTGTGGCTTCATCTACATACAGATCGCGCTTTTTCGCACACAGTCAG 176  
34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50  
|||||  
177 AATATGCAGATCGCTTGTCTGCAATTCAAGCAGAAATCACTGTACA 226  
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67  
|||||  
227 AGTCAGACCCAGATGTCTTATTGACCAGATTAGCCAGCAGCAAGGAAGAT 276  
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84  
|||||  
277 AGTGTCTTGAAGACAATGAAGCGTCAGACCGAGGAGTGGCAGAGT 326  
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100  
|||||  
327 TAAGGCTCTTGTTCAGGATCTTGAAGTAAGGTCATAAAAAAGTTGCATC 376  
101 GlyAspValGlnMetProValAlaAlaValValValMetAlaCysSerAr 117  
|||||  
377 GGAATGTACAGATCGCAGTGGCTGTGTAGTTGTTATGGCTTGCATCG 426  
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134  
|||||  
427 GCCTGACTACCTGGAAAAAGACTATTAAATCCATCTTAAAAATACCAATAT 476  
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150  
|||||  
477 CTGTTCGCGCAAAATATCTCTTTTCATATCCAGGATGGATCACATCCT 526  
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167  
|||||  
527 GATGTTAGGAAGCTTGTCTTGAGCTATGATCAGCTGACCTATATGCAGCA 576  
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184  
|||||  
577 CTGGATTTTGAACCTGTGCATACCTGAAAGACCGAGGGAGCTGATTGCAT 626  
184 yTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200  
|||||  
627 ACTACAAAATTCACGTCTATTACAAGTGGCATTTGGATCAGCTGTTTTC 676  
201 LysHisAsnPheSerArgValIleIleLeuGluAspMetCulIleAl 217  
|||||  
677 AAGCATAATTTTAGCCGTGTTATCATACTAGAAAGATGATATGAAATTCG 726  
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArg 234  
| |||||

727 CCTGATTTTTTTGACTTTTTTTGAGCTGGAGCTACTCTCTTGACAGAG 776  
234 sPLySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250  
|||||  
777 ACAAGTCGATATGGCTATTTCTTGAATGACAAATGGACAAATGCAG 826  
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe 267  
|||||  
827 TTTGTCCAAGATCTTATGCTCTTACCGCTCTGATTTTTCCTCCGGTCT 876  
267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrp 284  
|||||  
877 TGGATGGATGCTTTCAAAATCTACTTTGGGACGAATATCTCCAAATGGC 926  
284 rOLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300  
|||||  
927 CAAAGCTTACTGGGACGACTGGCTAAGACTCAAGAGAATCACAGAGT 976  
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluH 317  
|||||  
977 CGACAATTTATTCGCCCAGAAAGTTTCAGATCATATAAATTTTGGTGAGCA 1026  
317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334  
|||||  
1027 TGGTCTAGTTTGGGCGAGTTTTCAGGAGTATCTTTGAGCCAAATTAAC 1076  
334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350  
|||||  
1077 TAAATGATGTCCAGGTTGATGGAAGTCAATGGACCTTAGTTACCTTTTG 1126  
351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367  
|||||  
1127 GAGGACAATTACGTGAACACACTTTGTGTGACTTGTGTTAAAAAGCTTAAGCC 1176  
367 oIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384  
|||||  
1177 CATCCATGGAGCTGATGCTGTTTGAAGCATTTAACATAGATGGTGATG 1226  
384 aIArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400  
|||||  
1227 TGGGTATTAGTACAGATCACTAGACTTTGAAGATATCGCACGGCAA 1276  
401 PheGlyIlePheGluGlnTrpLysAspGlyValProArgAlaAlaTyrLy 417  
|||||  
1277 TTTGGCATTTTGAAGAAATGGAAGGATGGTGTACCACGGGCAGCATATA 1326  
417 sGlyIleValValPheArgPheGlnThrSerArgArgValPheLeuValS 434  
:::|||||  
1327 AGCAATAGTGGTTTCCGGTACCAACGTCAGACGCTGATTCCTTGTGTG 1376  
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446  
:::|||||  
1377 GCCTGATTGCTTCAACAACTCGGAAATGAAGATACT 1414

seq\_name: gb\_pi:NTA249883

seq\_documentation\_block:  
LOCUS NTA249883 1708 bp mRNA linear PLN 04-JAN-2002  
DEFINITION Nicotiana tabacum mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A9.  
ACCESSION AJ249883  
VERSION AJ249883.1 GI:18076147  
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.  
SOURCE common tobacco.  
ORGANISM Nicotiana tabacum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Nicotiana.  
1 (sites)  
Wunderoth, I., Tjaden, J. and von Schaeuwen, A.  
Isolation and characterization of different plant  
N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional  
analyses in the Arabidopsis cgl mutant, and in potato and tobacco

antisense plants  
Unpublished  
2 (bases 1 to 1708)  
AUTHORS von Schaeuwen, A.  
TITLE Direct Submission  
JOURNAL Submitted (28-SEP-1999) von Schaeuwen A., Pflanzenphysiologie,  
Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,  
GERMANY  
FEATURES  
Location/Qualifiers  
1..1708  
/organism="Nicotiana tabacum"  
/variety="Samsun NN"  
/db\_xref="taxon:4097"  
/clone="A9"  
/tissue\_type="leaf mesophyll"  
/note="sink"  
113..1453  
/gene="GntI"  
113..1453  
/gene="GntI"  
/EC\_number="2.4.1.101"  
/function="initiates complex-type glycans in the Golgi"  
/codon\_start=1  
/product="N-acetylglucosaminyltransferase I"  
/protein\_id="CAC80702.1"  
/db\_xref="GI:18076148"  
/translation="MRGNKFCDFRYLLILAAVFIYTMRLFATQSEYADRLAAAE  
AENHCTQTRLLIDQISLQOGRIVALEQMKRQDOECRLRALVODLESKGIKKLIGN  
VQMPVAAYVWACHNRADYLEKTIKSLKYQISVASKYPLFIQDGSHPDVRKLALSYD  
QLTYMQHLDPFEPHTERPELITATYKIAHRYKAWLDQLFYKHNSRVIILEDMDIAP  
DFFDFEAGATLLDRDKSISMAISSNNDNGOMQFVDDPYALYRSDFPGLGWLMSKSTW  
DELSKPWPKAYWDDMLRKENHGRQFIRPEVCRITYNFEHSGSSLGQFFKQYLEPIKL  
NDVQDWKSMDSLYLLEDNYVKFGDLVKAKPIHGADAVLKAFNIDGVRIOYRDQL  
DFENIAFOGIFEEWKDGVPRAAVKGIIVFVRYQTSRRRVFLVGHDSLQQLGIETD"  
BASE COUNT 506 a 326 c 380 g 496 t  
ORIGIN  
alignment\_scores:  
Quality: 2198.00 Length: 446  
Ratio: 5.030 Gaps: 0  
Percent Similarity: 97.982 Percent Identity: 92.825  
alignment\_block:  
US-09-591-466C-2 x NTA249883 ..  
Align seg 1/1 to: NTA249883 from: 1 to: 1708  
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17  
|||||  
113 ATGAGAGGGAACAAGTTTGTCTGTGATTTCCGGTACCTCCTCATCTTGGC 162  
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34  
|||||  
163 TGCCTGTCGCTTCATCTACACAGATCGGCTTTTTCGCACACATCAG 212  
34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50  
|||||  
213 AATATGCAGATCGCTTGTCTGTGCAATTTGAAGCAGAAAAATCATTTGTA 262  
51 SerGlnThrArgLeuLeuLeaspLysIleSerGlnGlnGlnGlyArgVa 67  
|||||  
263 AGCCAGACAGATTTGCTTATTGACCAGATTAGCCTGCACGAAGAGAAT 312  
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGlnCysArgGlnL 84  
:|||||  
313 AGTTGCTCTTGAAGAACAAATGAAGCGTCAGGACGAGGATGCCGACAAT 362  
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100  
|||||  
363 TAAGGCTCTTGTTCAGGATCTTGAAGTAAGGGCATAAAAAGTTGATC 412  
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117

[illegible]

```

1313 TTGTGGCATTTTGAAGAAATGGAAGATGGTGTCACCACTGCAGCATATAA 1362
      417 sGLyIleValValPheArgPheGlnThrSerArgValPheLeuVals 434
          |||||||.....:|||||:::|||||:::|||||:::
      1363 AGGAATAGTAGTTTCCGCGTACCACAACGTCAGACGGTATTCTCTGTG 1412
      434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
          :: |||||||:::|||||:::|||||:::|||||:::
      1413 GCCATGATCGCTTCAACAACACTCGGAATTGAAGATACT 1450

seq_name: gb_pat:A95055
seq_documentation_block:
LOCUS       A95055                      1737 bp    DNA     linear     PAT 26-JAN-2000
DEFINITION   Sequence 3 from Patent WO9929879.
ACCESSION   A95055
VERSION     A95055.1 GI:6779208
KEYWORDS    .
SOURCE      common tobacco.
ORGANISM    Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE   1 (bases 1 to 1737)
AUTHORS     Von,S.A.
TITLE       VEGETABLE GntI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH
            A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GntI)
ACTIVITY
JOURNAL     Patent: WO 9929879-A 3 17-JUN-1999;
FEATURES     VON SCHAEWEN ANTJE (DE)
             Location/Qualifiers
               source                1..1737
               /organism="Nicotiana tabacum"
               /strain="SAMSON NN"
               /db_xref="taxon:4097"
               /cell_type="BLATTZELLEN"
               /tissue_type="MESOPHYLL"
               /clone_lib="LAMBDA ZAP II (ECORI)"
               /clone_lib="(B) CLON(E); GNTI-A9(T)"
               /dev_stage="M; SINK' ORGAN"
               15..126
               127..1467
               /gene="CGL"
               127..1467
               /gene="CGL"
               /standard_name="GNTI"
               /EC_number="2.4.1.101"
               /function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.
               GLIKOPROTEINEN"
               /note="ERSTE GNTI-SEQUENZ AUS TABAK (UNPUBLIZIERT)"
               /codon_start=1
               /evidence=experimental
               /product="BETA-1,2-N-ACETYLGALUCOSAMINYLTRANSFERASE I"
               /protein_id="CAB70464.1"
               /db_xref="GI:6779209"
               /translation="MRGNKFCDDFRYLILAAVAFIYTQMLRFATQSEYADRLAAAI
               AENHCTSOTRLIDIQISQQFRRYLEEQMKRQDQCRLRALVDQLESKGIKKLISY
               VOMPYAAVVACNWRADLYETIKSLKYOISVASKYPLFI SQDGSHPDVKRLALSYD
               QLTMYHQILDFFEPHTRPECELTYAVKIARHYKWALDQLFYKNFSVIILEDMEIAP
               DFDFFEAGATLLDRKSTIMASSNDNCQMCFQVDYPALYSDFPPGLGWMLSKSTW
               DELSPKPWAYHDDLRLKENIRHGROFIRPEVCRTINFGEHSSLGQFFPKYLEPIKL
               NDVOQDWKSMDLSYLEDNIYNKHFGDLVKKAKPIHGADAVLKAFINDIGDVRIQYRDOL
               DFENIARQGFIFEEWKGDVPRAAYKGIIVVFYRTYRRVFLVGHDLSLQQLGIEDT"
               154..>213
               /gene="CGL"
               /standard_name="MEMBRANKER EINES TYP II GOLGI-PROTEINS"
               /function="MEMRANKER (AMINO ACID 10-29)"
               /note="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN
               IDENTIFIZIERT"
               /codon_start=1
               /product="HYDROPHOBES AMINO ACIDBEREICH IN GNTI"
               /protein_id="CAB70465.1"
               /db_xref="GI:6779210"

CDS

```



3'UTR /translation="FRYLILAAVAFIYQMRLF"  
BASE COUNT 510 a 337 c 390 g 500 t  
ORIGIN  
  
alignment\_scores:  
Quality: 2198.00 Length: 446  
Ratio: 5.030 Gaps: 0  
Percent Similarity: 97.982 Percent Identity: 92.825  
  
alignment\_block:  
US-09-591-466c-2 x A95055 ..  
  
Align seg 1/1 to: A95055 from: 1 to: 1737

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValAla 17  
127 ATGAGAGGAACAAGTTTGTGTGTTCCGGTACCTCCCTCATCTTGGC 176  
  
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34  
177 TGCTGTGCCCTTCATCTACACACAGATCGCGCTTTTTCGCACACAGTCAG 226  
  
34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50  
227 ANTATGCAGATCGCTTCTGCTGCAATTTGAAGCAGAAAATCATTTGTACA 276  
  
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlyArgVa 67  
277 AGCCAGACCCAGATTGCTTATTGACCAGATTAGCCTGCAGCAAGGAAGAA 326  
  
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84  
327 AGTTGCTCTTGAAGAACAAATGAAGCGTCAGGACCGAGGTGCGGCAAT 376  
  
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100  
377 TAAGGGCTCTTGTTCAGGACTCTTGAAGAGTAAGGCGATATAAAAGTTGATC 426  
  
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117  
427 GGAATGTACAGATGCCAGTGGCTGTGTAGTTGTTATGGCTTGCATCG 476  
  
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThr 134  
477 GGCTGATTACCTCGAAAGACTATTAAATCCATCTTAAATATCAAAATAT 526  
  
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150  
527 CTGTGGCTGCAAAATATCTCTTTTCATATCCAGGATGCCAGTGCATCCT 576  
  
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167  
577 GATGTCAGGAAGCTGCTTTGAGCTATGATCAGCTACGCTATATGCAGCA 626  
  
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184  
627 CTTGGATTTTGAACCTGTGCATACTCAAGAGACCAGGGAGCTGATTGCAT 676  
  
184 yTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200  
677 ACTACAAAATTTGCAGCTCATTAAGTGGGCTATGGATCAGCTGTTTAC 726  
  
201 LysHisAsnPheSerArgValIleIleLeuGluAspMetGluIleAl 217  
727 AAGCAATAATTTTACCGCTGTATCATACTAGAAGATGATGCAAAATGCC 776  
  
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArg 234  
777 CCTGATTTTTTGACTTTTTCAGGCTGGAGCTACTCTTCTTGACAGAG 826  
  
234 sPlysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250

||||| 827 ACAAGTCGATTATGCGCTATTTCTTCTTGGAAATGACAAATGGCAATGCAG 876  
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe 267  
877 TTTGTCCAAGATCCTTATGCTCTTTACCGCTCAGATTTTTCCTCCGGTCT 926  
267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrp 284  
927 TGGATGGATGCTTTCAAAATCTACTTGGGACGAATTTATCCAAAGTGGC 976  
284 roLysAlaIleTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300  
977 CAAAGGCTTACTGGGACGACTGGCTAAGACTCAAGAGAGATCACAGAGT 1026  
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHi 317  
1027 CGACAATTTATTCGCCCAGAGTTTGCAGAACATATAATTTTGTGTAGCA 1076  
317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334  
1077 TGGTCTAGTTTGGGCGAGTTTTCAGACAGTATCTTGAGCCCAATTAAC 1126  
334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350  
1127 TAAATGATGTCAGGTTGATTGGAAGTCAATGGACCTTAGTTACCTTTTG 1176  
351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367  
1177 GAGGACAAATTACGCTGAACACTTTGGTGACTTTGGTTAAAAGGCTAAGCC 1226  
367 olleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384  
1227 CATCCATGGAGCTGATGCTGCTTGAAGCATTTAACATAGATGGTGATG 1276  
384 aArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400  
1277 TGGTATTTCAGTACAGAGATCAACTAGACTTTGAAAATATATCGCACGCCAA 1326  
401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLy 417  
1327 TTTGGCATTTTGAAGAATGGAAGGATGGTGATCCACCGTCCACCATATA 1376  
417 sGlyIleValValPheArgPheGlnThrSerArgArgValPheLeuVal 434  
1377 AGGAATAGTAGTTTTCGGTACCAACGTCACGACGCTGATTCTCTGTTG 1426  
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446  
1427 GCCATGATTCGCTTCAACAACCTCGGAATTTGAAGATACT 1464  
  
seq\_name: gb\_pl:NTA249882

seq\_documentation\_block:  
LOCUS NTA249882 1836 bp mRNA linear PLN 04-JAN-2002  
DEFINITION Nicotiana tabacum mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A4.  
ACCESSION AJ249882  
VERSION AJ249882.1 GI:18076145  
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.  
SOURCE common tobacco.  
ORGANISM Nicotiana tabacum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
1 (sites)  
Wenderoth, I., Tjaden, J., and von Schaewen, A.  
Isolation and characterization of different plant N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1836)



|||||  
1277 AGGAATAGTGGTTTCGGTACCAACGTCACAGCTGATTCCCTGTGTTG 1326  
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446  
:|||||  
1327 GCCCTGATTCGCTTCAACAACACTCGGAATGAGATACT 1364  
seq\_name: gb\_pl:STU249880

seq\_documentation\_block:  
LOCUS STU249880 1499 bp mRNA linear PLN 04-JAN-2002  
DEFINITION Solanum tuberosum mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A8.  
ACCESSION AJ249880  
VERSION AJ249880.1 GI:18076143  
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (sites)  
Wenderoth, I., Tjaden, J. and von Schaewen, A.  
Isolation and characterization of different plant N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants  
Unpublished  
2 (bases 1 to 1499)  
von Schaewen, A.  
Direct Submission  
Submitted (28-SEP-1999) by Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY

FEATURES  
Source Location/Qualifiers  
1..1499  
/organism="Solanum tuberosum"  
/variety="Desiree"  
/db\_xref="taxon:4113"  
/clone="A8"  
/tissue\_type="tuber"  
/note="sink"  
37..1158  
/gene="GntI"  
37..1158  
/gene="GntI"  
/EC\_number="2.4.1.101"  
/function="initiates complex-type glycans in the Golgi"  
/codon\_start=1  
/product="N-acetylglucosaminyltransferase I"  
/protein\_id="CAC80699.1"  
/db\_xref="GI:18076144"  
/translation="MKRQDQECRLRALVODLESKGIKKLIGDVQMPVAVVVWACSR  
ADYLERTIKSLKYQTSVASKYPLFTSQDSNPDVRKLAISYDQLTVMQHLDFEPVHT  
ERPELVAYVKIARHYKALDQLFHKNSRVIILEDMEIAADFTDYFAGATLLDR  
DKSLIMJSSNDNGQKQFVQDPDALYNSDFPGLGWLMSKSTWSELSPKWPAYWDDW  
LRUKENHRGQFJHPEVCRTYFEGHSSLGQFFKQYLEP IKLNDVQVQWKSMDLSYL  
LEDNVKHFQDGLVKKAKPHGADAVLKAFNIDGVDVRIQVRLDQDFEAIARQGIPEEW  
KDGVPRAAYKGIWVFRFQTSRRVFLVAPDSLRLQGVEDT"  
BASE COUNT 452 a 266 c 339 g 442 t  
ORIGIN

alignment\_scores:  
Quality: 1948.00 Length: 374  
Ratio: 5.237 Gaps: 0  
Percent Similarity: 99.465 Percent Identity: 98.128

alignment\_block:  
US-09-591-466C-2 x STU249880 ..

Align seg 1/1 to: STU249880 from: 1 to: 1499

73 GlnMetLysHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValG1 89  
|||||  
34 CAATGAAGCGCTCAGGACGAGCTGCGGCAATTAAGGCTCTGTGTCA 83  
89 nAspLeuGluSerLysGlyIleLysLysLeuIleGlyAspValGlnMetP 106  
|||||  
84 GGATCTTGAAGTAAAGGCATAAAAAGTTAATCGGAGATGTCAGATGC 133  
106 roValAlaAlaValValMetAlaCysSerArgThrAspTyrLeuGlu 122  
|||||  
134 CAGTGGCAGCTGTAGTTGTTATGCTTGCAGTCGCTGACTACCTGGAG 183  
123 ArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaSerLysTy 139  
|||||  
184 AGGACTATTAAATCCATCTTAAATACCAACATCTCTTGCATCAAAATA 233  
139 rProLeuPheIleSerGlnAspGlySerAsnProAspValArgLysLeuA 156  
|||||  
234 TCCTCTTTTCATATCCAGGATGGATCAATCTGTATGTAAGAAGCTTG 283  
156 laLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluPro 172  
|||||  
284 CTTTGAGCTATGATCAGCTGACGTATATGACGACCTTGGATTTGAACCT 333  
173 ValHisThrGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaAr 189  
|||||  
334 GTCCATACTGAAAGACGAGGGGAACCTGGTTGCATACATAAGATTGCACG 383  
189 gHisTyrLysTyrPalaLeuAspGlnLeuPheHisLysHisAsnPheSerA 206  
|||||  
384 TCATTACAAGTGGCATTTGCATCAGCTGTTTCCAAAGCATATTTTAGCC 433  
206 rgValIleIleLeuGluAspAspMetGluIleAlaAlaAspPhePheAsp 222  
|||||  
434 GTGTTATCATCTAGAAGATGATATGGAATTCGCTGCTGATTTTTCGAC 483  
223 TyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAl 239  
484 TATTTTGAGGCTGGAGCTACTCTCTTGCACAGACAACTCGATTATGGC 533  
239 aIleSerSerTyrAsnAspAsnGlyGlnArgGlnPheValGlnAspProA 256  
|||||  
534 TATTTCTTCTTGGAAATGCAATGGACAAAGGCACTTCCTCAAGATCCTG 583  
256 spAlaLeuTyrArgSerAspPheProGlyLeuGlyTyrMetLeuSer 272  
584 ATGCTCTTTACCGCTCAGACTTTTTCCTGCTTGGATGATGCTTTCA 633  
273 LysSerThrTyrSerGluLeuSerProLysTyrProLysAlaTyrTipAs 289  
634 AAATCAACTTGGTCCGAACCTATCTCCAAAGTGCCAAAGGCTTACTGGGA 683  
289 pAspTyrLeuArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgP 306  
684 TGACTGGCTAAGCTTGAAGAAAAATCACAGAGTTCGACAAATTTATTCA 733  
306 roGluValCysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGly 322  
734 CAGAAGTTTGCAGAACCTACAAATTTTGGTGAGCATGCTCTAGTTGGGG 783  
323 GlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnVa 339  
784 CAATTTTAAAGCAGTATCTTTGAGCCAAATTAAGCTAAATGATGCCAGGT 833  
339 lAspTyrLysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValL 356  
834 TGATTGGAAGTCAATGGACCTAAGTTACCTTTTGGAGGACAACTATGTGA 883  
356 yHisPheGlyAspLeuValLysLysAlaLysProIleHisGlyAlaAsp 372  
884 AACACTTTGGCGACTTGGTTAAAAAGGCTTAAGCCCATCCACGAGCTGAT 933  
373 AlaValLeuLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrAr 389

```

934 GCTGTTTTGAAGCATTATACATAGATGGTATGCGCTATTTCAGTACAG 983
389 gAspGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluG 406
984 AGACCAACTAGACTTTGAAGCTATCGCTCGACAGTTTGGCATTTTGAAG 1033
406 luTrpLysAspGlyValProArgAlaAlaTyrLysGlyIleValPhe 422
1034 AATGAAGATGGTGTACACCGGCAGCATATAAAGGATAGTAGTTTC 1083
423 ArgPheGlnThrSerArgValPheLeuValSerProAspSerLeuAr 439
1084 CGGTTTCAACATCTAGACGTGTCTCTCTGTTGCCCTGATTCTCTTCG 1133
439 gGlnGlyValGluAspThr 446
1134 ACAACTGGAGTTGAAGATACT 1155

seq_name: gb_pl:ATH243198
seq_documentation_block:
LOCUS ATH243198 1820 bp mRNA linear PLN 24-AUG-1999
DEFINITION Arabidopsis thaliana mRNA for N-acetylglucosaminyltransferase I
(GlcNAcT-I gene).
ACCESSION AJ243198
VERSION AJ243198.1 GI:5139334
KEYWORDS GlcNAcT-I gene; glycosyl transferase;
N-acetylglucosaminyltransferase I.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1820)
AUTHORS Bakker,H., Lommen,A., Jordi,W., Stiekema,W. and Bosch,D.
TITLE An arabidopsis thaliana cDNA complements the
N-acetylglucosaminyltransferase I deficiency of CHO lec1 cells
JOURNAL Biochem. Biophys. Res. Commun. 261 (3), 829-832 (1999)
MEDLINE 99373163
REFERENCE 2 (bases 1 to 1820)
AUTHORS Bakker,H.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Bakker H., Molecular Biology, Centre for
Plant Breeding and Reproduction Research, P.O. Box 16 Wageningen,
6700AA, NETHERLANDS
COMMENT Related sequence AL035538.
FEATURES
Location/Qualifiers
1..1820
/organism="Arabidopsis thaliana"
/variety="Colombia"
/db_xref="taxon:3702"
/clone_lib="pABE-ALGN1"
31..1365
/gene="GlcNAcT-I"
31..1365
/gene="GlcNAcT-I"
/EC_number="2.4.1.101"
/function="glycosyltransferase"
/codon_start=1
/product="N-acetylglucosaminyltransferase I"
/protein_id="CAB45521.1"
/db_xref="GI:5139335"
/translacion="MARISCDLRFLLIPAAFMFIYIQMLFOTOYADRLSSAIESE
NHCTSQMRGLIDEVSIKQSRVADLMKRDDELVOLKDIOTFEKKGIKALTQGGQ
MPVAAYVACSRADYLVKTQYTPVASKYPLFIQDGSQAVKSKLSYNOL
TYMQLHDFEVPVTERPGLTAYKIAHYKWLDDOLFYKHKSPVITLEDDEIAPDF
FDYFEAASLMDRDKTIMAASSNDNCKQFVHDVYALYSDFPGLGWLKKTWDE
LSPKWPAYWDLRLKENHKKQFIRPEVCRTYNFGEHSSLGQFESYLEPIKLD
VYDVKAKDLGLYTEGNTKIFSGLVQRAPRIQGSDLVLKAKNIKDDVIRYKDOVEF
ERAGEGFIPEEWKWDGVPRTAYKGVVVFYQITTRFVLVGPDSVMQLGRNS"
532 a 342 c 421 g 525 t
BASE COUNT

```

## ORIGIN

## alignment\_scores:

Quality: 1781.00 Length: 438

Ratio: 4.441 Gaps: 0

Percent Similarity: 91.553 Percent Identity: 75.571

## alignment\_block:

US-09-591-466c-2 x ATH243198 ..

Align seg 1/1 to: ATH243198 from: 1 to: 1820

```

9 AspLeuArgTyrLeuLeuValValAlaAlaLeuAlaPheIleTyrIleGl 25
49 GACTTGAGATTCTCTCATCCCGGACGTTTCATGTTCACTACATCCA 98
25 nMetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAlaAa 42
99 GATGAGGCTTTTCCAGACGCAATCACAGATATGCAGATCGCCTCAGTTCCG 148
42 laileGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAsp 58
149 CTATCGAATCTGAGAACCATTCAGTCAATGCGAGGCTCATAGAT 198
59 LysIleSerGlnGlnGlnArgValValAlaLeuGluGlnMetLy 75
199 GAAGTTAGCATCAACACAGTCGCGATTGTTGCCCTCGAAGATATGAAGAA 248
75 sHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValGlnAspLeuG 92
249 CCGCCAGACGAGAACTTGTGAGCTTAAGGATCTAATCCAGACGTTG 298
92 luSerLysGlyIleLysLysLeuIleGlyAspValGlnMetProValAla 108
299 AAAAAAAGAAATAGCAAACTCACTCAAGGTGACAGATGCGCTGTGGCT 348
109 AlaValValMetAlaCysSerArgThrAspTyrLeuGluArgThrI1 125
349 GCITGATGTTATGGCTCGAGTCGAGTCTGAGACTATCTTGAAGGACTGT 398
125 eLysSerIleLysTyrGlnThrSerValAlaSerLysTyrProLeuP 142
399 TAAATCAGTTTAAACATATCAAACTCCCGTTGCTTCAAAATATCCTCTAT 448
142 heileSerGlnAspGlySerAsnProAspValArgLysLeuAlaLeuSer 158
449 TTATATCTCAGGATGGATCTGTCAAGCTGTCAAGACCAAGTCATTGAGC 498
159 TyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluProValHisTh 175
499 TATAATCAATTAAACATATATGCAGCACTTGGATTTTGAACCCAGTGTCTAC 548
175 rGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrL 192
549 TGAAGGCGCTGCGCAACTGACTCGGTACTACAAGATTGCACGCTCACTACA 598
192 yStrAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIle 208
599 AGTGGCACTGGACCACTGTTTACAAACACAAATTTAGTCGAGTGATT 648
209 ileLeuGluAspAspMetGluIleAlaAlaAspPheAspTyrPheGl 225
649 ATACTAGAAGATGATGATGAAATTCCTCCAGACTCTTTTGATTACTTTGA 698
225 uAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSers 242
699 GGCTCGACCTACTCTCATGGATAGGATAAAAACCATTTATGGCTGCTTCAT 748
242 erTApAsnAspAsnGlnGlnArgGlnPheValGlnAspProAspAlaLeu 258
749 CATGGAATGATAATGCAGACAGACAGTGTTCATGATGATCCCTATCGCTTA 798

```

```
259  TyrArgSerAspPhePheProGlyLeuGlyTyrPmetLeuSerLysSerTh 275
799  TACCGATCAGATTTTCTCGCTTGGTGGATGCTCAAGAGATCGAC 848
275  rTrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpL 292
849  TTGGGATGAGTTATACCAAGATGGCCAAAGCCTTACTGGGATGATTGGC 898
292  euArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVal 308
899  TGACACTAAAGGAAAAACCATAAAGCCGCCCAATTCATTCGACCGGAAGTC 948
309  CysArgThrTyrAsnPheGlyLuhisGlySerSerLeuGlyGlnPhePh 325
949  TGTGAACATACAATTTTGGTGAACATGGGTCTAGTTTGGGACAGTTTTT 998
325  elysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrpL 342
999  CAGTCAGTATCTGGAACCTATAAAGCTAAACGATGTGACGGTTGACTGGA 1048
342  ysSerMetAspLeuSerTyrLeuGluAspAsnTyrValLysHisPhe 358
1049  AAGCAAGGACCTGGGATACCTGCAGAGGGAACATATACCAAGTACTTT 1098
359  GlyAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLe 375
1099  TCTGGCTTAGTCAGACAGCAGCAGCAATTCAGGTTCTGACCTTGTCTT 1148
375  ulysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAspGlnL 392
1149  AAAGGCTCAAAACATAAAGGATGATGTCGTATCCGGTATAAAGACCAAG 1198
392  euAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLys 408
1199  TAGAGTTTGAACCAATTCGACGGGAATTTGTATATTTGAAGAATTCGAAG 1248
409  AspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPheG 425
1249  GATGCTGTGCTCGCAACAGCATATAAAGGAGTACTGTTTTCGAATCCA 1298
425  nTrpSerArgValPheLeuValSerProAspSerLeuArgGlnLeuG 442
1299  GACAACAGCGTGTATTCCTGGTTGGGCCAGATTCTGTAATGCAGCTTG 1348
442  lyValGluAspThr 446
1349  GAATTCGAAATTC 1362
```

seq\_name: gb\_pl:ATH249881

seq\_documentation\_block:

LOCUS ATH249881 1830 bp mRNA linear PLN 04-JAN-2002  
DEFINITION Arabidopsis thaliana mRNA for N-acetylglucosaminyltransferase I (GntI gene).

ACCESSION AJ249881

VERSION AJ249881.1 GI:18072832

KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (sites)

Wenderoth,I., Tjaden,J. and von Schaewen,A.

Isolation and characterization of different plant

N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional

analyses in the Arabidopsis cgl mutant, and in potato and tobacco

antisense plants

Unpublished

REFERENCE 2 (bases 1 to 1830)

AUTHORS von Schaewen,A.

TITLE Direct Submission

JOURNAL Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie,

Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY

FEATURES

source

Location/Qualifiers

1..1830

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db\_xref="taxon:3702"

/tissue\_type="whole plant"

117..1451

/gene="GntI"

117..1451

/gene="GntI"

/EC\_number="2.4.1.101"

/function="initiates complex-type glycans in the Golgi"

/codon\_start=1

/product="N-acetylglucosaminyltransferase I"

/protein\_id="CAC80700.1"

/db\_xref="GI:18072833"

/translation="MARISCDLRFLLIPAAFMFYIQMLRFOTSOYADRLSSAIESE

NHCTSMRGLIDEVSIKQSRIVALEDMKRPDEELVOLKDLIQTFKKGIKLTGGO

MPVAAVVMACSRADYLVRTVSVLYQTPVASKYPLFISQDSDQAVKSKSLYNOL

TYMHLDPEFVTPERPELTATYTKIARHYKWLQDLFYHKFSRVIILEDDEIAPDF

FDYFEAAASLMDRDKTMAAASSNDNGQKQFVHDVYALRSDFPGLGMLKRSTWDE

LSPKPKAYDDMLRLKHKGRQFIAPVCRTYFNGHSSGSLGQFFSOYLEPIKLND

VTVDMKAKDLYLTGNTYTKYFSGLVQRPIQGSDLVLKAQNIKDDRIYKQDVEF

ERLAGEFGIFEWKQGVPRYAYKGVVVERIQITRRVFLVGPDSVMQLGRNS"

BASE COUNT 540 a 348 c 422 g 520 t

ORIGIN

alignment\_scores:

Quality: 1768.00 Length: 438

Ratio: 4.431 Gaps: 0

Percent Similarity: 91.096 Percent Identity: 75.114

alignment\_block:

US-09-591-466C-2 x ATH249881 ..

Align seg 1/1 to: ATH249881 from: 1 to: 1830

9 AspLeuArgTyrLeuLeuValValAlaLeuAlaPheIleTyrIleG 25

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

135 GACTTGAGATTCTTCATCCCGCAGCTTTCTATGTTCTACATCCA 184

25 nMetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAlaAla 42

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

185 GATGAGGCTTTTCAGACGCAATCACAGTATGCAGATCGCTCAGTTCCG 234

42 laIleGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAsp 58

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

235 CTATCGAATCTGAACACCATTCGACTAGTCAAAATGCGAGGCTCATAGAT 284

59 LysIleSerGlnGlnGlnArgValValAlaIleGluGlnMetLys 75

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

285 GAAAGTTAGCATCAACACAGTCGCGGATTTGTCCTCGAAGATATAGAA 334

75 sHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValGlnAspLeu 92

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

335 CCGCCAGGACGAAAGACTTGTGCGAGCTTAAGGATCTAATCCAGACGTTG 384

92 luSerLysGlyIleLysLysLeuIleGlyAspValGlnMetProValAla 108

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

385 AAAAAAAGGAATAGCAAAACTCACTCAAGGTGGACAGATGCTGTGGCT 434

109 AlaValValValMetAlaCysSerArgThrAspTyrLeuGluArgThrI 125

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

435 GCTGTAGTGGTTATGGCCCTGCAGTCGTGCAGACTATCTTTGAAAGGACTGT 484

125 eLysSerIleLeuLysTyrGlnThrSerValAlaSerLysTyrProLeu 142

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

485 TAAATCAGTTTAAACATATCAAACTCCCGTTGCTTCAAAATATCCTCTAT 534

142 heIleSerGlnAspGlySerAsnProAspValArgLysLeuAlaLeuSer 158

```
|||||
535 TTATATCTCAGATGGATCTGATCAAGCTGTCAAGACGACGATCAATTGAGC 584
|||||
159 TyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluProValHisTh 175
|||||
585 TATAATCAATTAACATATATGACGACCTTGGATTTTGAACGAGTGGTCAC 634
|||||
175 rGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrL 192
|||||
635 TGAAGGCCTGCTGAAGTACTGCTGCTACTACAAGATTGCACGTCACAC 684
|||||
192 ySTrpAlaLeuAspGlnLeuPheHisTyrHisAsnPheSerArgValIle 208
|||||
685 AGTGGCAGCTGCAGCAGTGTGTTTACAAACACAAATTTAGTCGAGTGATT 734
|||||
209 IleLeuGluAspAspMetGluIleAlaAlaAspPheAspTyrPheG1 225
|||||
735 ATACTAGAACGATATGGAAATGTCTCCAGACTTCTTTGATTACTTTGA 784
|||||
225 uAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSerS 242
|||||
785 GCGTGCAGCTAGTCTCATGATAGGATAGGATAAACCATTATGCGTCTCAT 834
|||||
242 erTrpAsnAspAsnGlyClnArgClnPheValGlnAspProAspAlaLeu 258
|||||
835 CATGGAATGATAATGGACAGAAGCAGTTTGTGCATGATCCCTATGCGCTA 884
|||||
259 TyrArgSerAspPheProGlyLeuGlyTyrPheMetLeuSerLysSerTh 275
|||||
885 TACCGATCACATTTTCTCGCCTTGGGTGGATGCTCAAGAGATCGAC 934
|||||
275 rTrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpL 292
|||||
935 TTGGGATGAGTTATCACCAAGTGGCCAAAGGCTTACTGGGATGATTGC 984
|||||
292 euArgLeuLysGluAsnHisArgGlyArgClnPheIleArgProGluVal 308
|||||
985 TGAGACTAAGGAAACCATATAAGCGCGCAATTCATTGCACCGGAAGTC 1034
|||||
309 CysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyClnPheP 325
|||||
1035 TGPAGAACATACAAATTTTGTGAACATGGTCTAGTTTGGGACAGTTT 1084
|||||
325 eLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrpL 342
|||||
1085 CAGTCAGTATCTGGAACCTATAAAGCTAAACGATGTGACGGTTGACTG 1134
|||||
342 ySerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPhe 358
|||||
1135 AAGCAAAGCACCTGGCATACCTGCAGAGGGAACCTATACCAAGTACTTT 1184
|||||
359 GlyAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaVal 375
|||||
1185 TCTGGCTTAGTCAGACAAACAGACCAATTCAGGTTCTGACCTTGTCTT 1234
|||||
375 uLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAspGlnL 392
|||||
1235 AAAGCTCAAAACATAAAGGATGATGCTATCCGATAAAGACCAAG 1284
|||||
392 euAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGlnTrpLys 408
|||||
1285 TAGAGTTTGACGCATTTGAGGGAATTTGGTATATTTGAAGAATGGAAG 1334
|||||
409 AspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPheG1 425
|||||
1335 GATGTTGCCACGACGACATATAAAGGAGTAGTGGTGTTCGAATCCA 1384
|||||
425 nThrSerArgArgValPheLeuValSerProAspSerLeuArgGlnLeuG 442
|||||
1385 GACAACAAGACGTGATTCTCTGGTTGGGCCACAGATTCGTATATGACGCTG 1434
|||||
442 lyValGluAspThr 446
|||||
```

1435 GAATTCGAATTC 1448

seq\_name: gb\_pat:A95057

seq\_documentation\_block:

LOCUS A95057 1854 bp DNA linear PAT 26-JAN-2000  
DEFINITION Sequence 5 from Patent WO9929879.  
ACCESSION A95057  
VERSION A95057.1 GI:6779211

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 1854)

Von, S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH

A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GNTI)

ACTIVITY

Patient: WO 9929879-A 5 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1. 1854

/organism="Arabidopsis thaliana"

/strain="COLUMBIA"

/db\_xref="taxon:3702"

/tissue\_type="ALLE GEWEBE"

/clone\_lib="LAMBDA UNI-ZAP (ECORI/XHOI) UND LAMBDA ACT"

/clone\_lib="{XHOI}"

/dev\_stage="M: REIFE PFLANZEN"

19. .134

135. .1469

/gene="CGL"

/gene="CGL"

/standard\_name="GNTI"

/EC\_number="2.4.1.101"

/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET."

GLYKOPROTEINEN"

/note="ERSTE GNTI-SEQUENZ AUS ARABIDOPSIS (UNPUBLIZIERT)"

/codon\_start=1

/evidence=experimental

/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I"

/protein\_id="CAB70466.1"

/db\_xref="GI:6779212"

/translation="MARISCDRLFLIPAAFMFIYQMRLFTQTSOYADRLSSAIESE

NHCTSQMRGLIDEVSIKQSRIVALEDMKNRQDEELVOLKDIOTFEKKGTAKLTQGSQ

MPVAAVVMACSRADYLVRTVKSVLTVTPVASKYPLFISODGSDQAVKSKSYNQL

TYMHLDPEPVTERGELTAYKYIARHYKWLQDLFYKHKFSRVILEDDMETAPDF

FDYFEAAASLMRDKTI MAASSWNNGKQFVHDHPYALYKRSDFPGLGMLKSTWDE

LSPKPKAYWDDMLRLKENHKHGFIAPEVCRTYNGEHSIGQFFSQYLEPIKLN

VTVDKAKGLYTEGNTYKFSGLVRQARPIQGSOLVLKAQNIKDDIRIRYKQVEF

ERIAEGFGIFEEMKDGVPRTAYKGVVVFRIQTTRRVFLVGPDSVMQLGRNS"

<157..>215

/gene="CGL"

/standard\_name="MEMBRANKER EINES TYP II GOLGI-PROTEINS"

/function="MEMBRANKER (AMINO ACID 8-27)"

/note="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN

IDENTIFIZIERT"

/codon\_start=2

/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"

/protein\_id="CAB70467.1"

/db\_xref="GI:6779213"

/translation="EISSHPGSHVHLHPDEAF"

1470..1848

3' UTR 545 a 357 c 430 g 522 t

BASE COUNT  
ORIGIN

alignment\_scores:

Quality: 1768.00

Length: 438

```

Ratio: 4.431      Gaps: 0
Percent Similarity: 91.096      Percent Identity: 75.114

alignment_block:
US-09-591-466C-2 x A95057      ..

Align seg 1/1 to: A95057 from: 1 to: 1854

9  AspLeuArgTyrLeuLeuValValAlaAlaLeuAlaPheIleTyrIleG1 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 GACTTGAGATTCTTCATCCCGGACGCTTCATGTTTCATCTACATCA 202
25  nMetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAlaAla 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 GATGAGGCTTTTCAGACGCAATCACAGTATGCGATCGCTCGATTCCG 252
42  laIleGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAsp 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 CTATCGAATCTGACAACCATGTGCACTAGTCAAAATGGCGGCGCTCAT 302
59  LysIleSerGlnGlnGlnArgValValAlaLeuGluGluGluMetIly 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 GAAGTTAGCATCAACAGTCCGGATTGTTGCCCTCAAGATATGAAGA 352
75  sHiGlnAspGlnGlnCysArgGlnLeuArgAlaLeuValGlnAspLeuG 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 CCGCAGGACGCAACAACTGTGCGAGCTTAAGGATCTAATCCAGACGCT 402
92  luSerLysGlyIleLysLysLeuIleClyAspValGlnMetProValAla 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 AAAAAAAGGAATAGCAAACTCCTCAAGGTGGACAGATGCTGTGGCT 452
109 AlaValValMetAlaCysSerArgThrAspTyrLeuGluArgThrI1 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453 GCTGTAGTGTTATGGCTCGCAGTCGTCGACACTATCTTGAAGGACTCT 502
125 eLysSerIleLeuLysTyrGlnThrSerValAlaSerLysTyrProLeuP 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
503 TAAATCATGTTTAAACATATCAAACTCCGTTGCTTCAAAATATCCTCT 552
142 helLeSerGlnAspClySerAsnProAspValArgLysLeuAlaLeuSer 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
553 TTATATCTCAGATGGATCTGATCAAGCTGCTCAAGACCAAGTCAATG 602
159 TyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluProValHis 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
603 TATATCAATTAAGATATATGCAAGCACTTGGATTTTGAACCAAGTGT 652
175 rGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyr 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
653 TGAAGGCCCTGGTCACTGACTCGGTACTACAACATGTCACGTCACTAC 702
192 ysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIle 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
703 AGTGGGCACCTGGACAGCTGTTTTACAACACACAAATTTAGTCGAGTG 752
209 lLeuGluAspAspMetGluIleAlaAlaAspPheAspTyrPheG1 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
753 ATACTAGAAGCAGATATGGAATTTGCTCCAGACTTCTTGATTACTTTGA 802
225 uAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSer 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
803 GGCTGCAGCTAGTCTCATGATAGGATAAAACCAATTATGGCTGCTTCAT 852
242 erTrpAsnAspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeu 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
853 CATGGAATGATTAATGGACAGAAGCAGTTTGTGTCATGATCCCTATGCG 902
259 TyrArgSerAspPheProGlyLeuGlyTyrPheLeuSerLysSerThr 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
903 TACCATCATGATTTTTTCTTGCCCTTGGGTGATGCTCAAGAGATCGAC 952
275 rTrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpL 292

```

953	TTGGGATGAGTTATCACCAGATGGCCAAAGGCTTACTGGATGATTGGC	1002
292	euArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVal	308
1003	TCAGACTAAGGAAACCAATAAAGGCGCCAATTTCATTCACCGGAATC	1052
309	CysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPheP	325
1053	TCTAGAACATACAAATTTGGTGAACATGGGTCTAGTTTGGACAGATTTT	1102
325	elysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrpL	342
1103	CAGTCAGTATCTGGAACTTAAAGCTAAACGATGTGACGTTGACTGGA	1152
342	ysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPhe	358
1153	AAGCAAGACGCTGGGATACCTCACAGAGGGAAACTATACCAAGTACTTT	1202
359	GlyAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValle	375
1203	CTCGGCTTAGTCAGACACGACGACCAATTCAAGGTTCTGACCTTGCTCT	1252
375	uLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAspGlnL	392
1253	AAAGGCTCAAAACATAAAGGATGATCGTATCCGCTATAAAGACCAAG	1302
392	euAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLys	408
1303	TAGAGTTTGAACGCATTCGAGGGGAATTTGGTATATTTGAAGAATGGAAG	1352
409	AspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPheGl	425
1353	GATGGTGTCCACGACACCATATAAAGCAGTAGTGGTCTTCGAATCCA	1402
425	nThrSerArgArgValPheLeuValSerProAspSerLeuArgGlnLeuG	442
1403	GACACAGACGCTGATTCTCTGGTTGGGCCAGATTCTGTAATGCAGCTTG	1452
442	lyValGluAspThr	446
1453	GAATTCGAAATTC	1466
seq_name: gb_pl:ATF20D10		
seq_documentation_block:		
LOCUS	ATF20D10	147215 bp DNA linear
DEFINITION	Arabidopsis thaliana DNA chromosome 4, BAC clone F20D10 (project).	
ACCESSION	AL035538.1	GI:4467094
VERSION	AL035538	
KEYWORDS		
SOURCE		
ORGANISM	Arabidopsis thaliana	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; Brassicales; Brassicaceae; Arabidopsi	
REFERENCE	1 (bases 1 to 147215)	
AUTHORS	Bevan, M., Wedler, H., Kutzner, M., Wambutt, R., Bancroft, M., Mayer, K.F.X. and Schueller, C.	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 147215)	
AUTHORS	EU Arabidopsis sequencing project.	
TITLE	Direct Submission	
JOURNAL	Submitted (26-FEB-1999) MIPS, at the Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de	
	Coordinator: Mike Bevan, Molecular Genetics Department, Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK	
	E-mail: michael.bevan@jic.ac.uk	
	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosome 4 is available at: <a href="http://www.mips.biochem.mpg.de/proc/thaliana/">http://www.mips.biochem.mpg.de/proc/thaliana/</a>	
COMMENT		

FEATURES	source	Location/Qualifiers
		1..147215
		/organism="Arabidopsis thaliana"
		/variety="Columbia"
		/db_xref="taxon:3702"
		/chromosome="4"
gene		complement(join(830..2566,2624..3049))
gene		/gene="F20D10.10"
		830..3049
		/gene="F20D10.10"
CDS		complement(join(830..2566,2624..3049))
		/gene="F20D10.10"
		/note="strong similarity to retrotransposon -like protein
		-Arabidopsis thaliana, PID:e1248853"
		/codon_start=1
		/product="putative protein"
		/protein_id="CAB37529.1"
		/db_xref="GI:4467095"
		/translation="MMNGLRRTFWSSIHKKKDNNRVDDSLDRQKPTTTSRRGFFSNPS TPRSETPDSVTCSPITPCQSWSATAIPTSPSLPASPKLQCDTSGDVTPTNRSPLS FLSVSSSSSTPSPKSPASFSLKSLCKFTKVRTLNRRNQSSISSKCGICLOSAGK RGTAIFPACSHTFHFPCVASRAGDRNLSDPCVGASWRETSLPLSLSSLHESGS ESDSKIRESKNNKSLRVYNDDEPLISSPISRTGFTIPESNEDEEEDNDGEGKGF YNTPLSTKMLTDSVTGHVDKLVLSSEAAIVAVGRNETYSVLWKIKSPSLPTARR SPVDLVTVDVSGGNIEMVKRAMROVSISSRETDRLSMVSFSSSKRLTPLRMTANG RLRLARRIVDDISDGDGMSVNDVAKKAIVIEDRQKNLFTTIFVLTDRNNSAHQAQ LAQDPVTFSTRFSLHLEIPHTITWLGACNHLPEDVFAKRIKLSLSVODLTLNLGLV SGSGGQKVTSVLSGRPVWLSGLIRGDMYGDDEEREVLVELKSPSSRSQRIMTVR SHRVDPTQEIKNYGDRLMTIPPTTVRSSDPSIARLNLHVSTRAVAESRRLVEVND YSGAERMARALLVQGLSSDCLRGLEAELADLNLRLGRHVAVSPSPVQKSE PLTPTSAMRAAERLAKVAITMRKHMNRVSLHGFENAR"
exon		complement(830..2566)
		/gene="F20D10.10"
		/number=1
intron		complement(2567..2623)
		/gene="F20D10.10"
		/number=1
exon		complement(2624..3049)
		/gene="F20D10.10"
		/number=2
repeat_region		5063..5098
exon		/note="35bp ta tandem repeat"
		complement(9755..10264)
		/gene="F20D10.20"
		/number=1
gene		9755..12463
		/gene="F20D10.20"
gene		complement(join(9755..10264,10338..11657,11741..11989, 12071..12191,12300..12463))
		/gene="F20D10.20"
CDS		complement(join(9755..10264,10338..11657,11741..11989, 12071..12191,12300..12463))
		/gene="F20D10.20"
		/note="similarity to Caenorhabditis elegans cosmid F32B5, gene F32B5.7, PID:g2088783
		Contains Eukaryotic putative RNA-binding region RNP-1 signature [KGSCLPDM]
		contains EST gb:N97112:AA728688"
		/codon_start=1
		/product="putative protein"
		/protein_id="CAB37530.1"
		/db_xref="GI:4467096"
		/translation="MDKEEQTLWEAQAQKIDISVLLAAAKKHLFLGAVDRNRCILY DGPALQRAIYRYNAYWLPILQAQTESSICQGLVPPLDCEWYHCHRLNPVRYKTDG EQFGRLVDNSGVSSVNGKSKQETLWKRLYPTEPYDLDFANAISEPADVSALSKC TYDVLAVLRQSPFYQVQSAHVNDVFLQEAARYKAFYLIKGNRRS IKLCVVP TYDILWHTHQLHAIYSYCNDLTKMIGKVLHDDTSDRSKGLDTSFGSITQAQHEE TFGRYVWKACAMKNGMTPKPVTTPYVYCVSKKSIKAEESQVNIQIPKRVHLEILEI VGKNLDPDAHKGVVFLSKTQPSLFPNERRLTVLSESCGEKQVAFQCEPTGELSF QLMSSKSLGFTSLFSBFLSPVTKLSVEKMLELPTTKRGKADDPNPISLRVAVSFT PPTRSPVTLHVAQRPVSLKSGCFLPMLRVLAQSTRYVDETEVI NLQWRNSDA APGRDRQVIGVKECGEYVLAIEDGTFWSLDKWSLQTCNPATDGPLFELSGTRM VKVYSRKLKLEYPKHCILSRSEQDFMTAVEFSKQHPYKAVGLGLDLKFGSIEANEKWL

VLPQVMSFILLSLLKKEGFSAAKQDTVKANGITEESTIDVLSQEKLEETNMDDVDT  
TTPVAVAAEKINGGARCFSEKLSGNMIEEGGHCAGCGGCGGGGGGGGRCGGM  
TKLEGCGGSGCTGGSTGCNCGCGGCGGMMKNANNANGNAPSVENDAVTA"

complement(10265..10337)

/gene="F20D10.20"

/number=1

complement(10338..11657)

/gene="F20D10.20"

/number=2

complement(11658..11740)

/gene="F20D10.20"

/number=2

complement(11741..11989)

/gene="F20D10.20"

/number=3

complement(11990..12070)

/gene="F20D10.20"

/number=3

complement(12071..12191)

/gene="F20D10.20"

/number=4

complement(12192..12299)

/gene="F20D10.20"

/number=4

complement(12300..12463)

/gene="F20D10.20"

/number=5

complement(join(13386..14030,14118..14738,14828..15394,  
15491..15569,15669..15730,16043..16069))

/gene="F20D10.30"

13386..16069

/gene="F20D10.30"

complement(join(13386..14030,14118..14738,14828..15394,  
15491..15569,15669..15730,16043..16069))

/gene="F20D10.30"

/note="Strong similarity to dnaK-type molecular chaperone  
PHSP1 -Pisum sativum, PID:g20835

Contains ATP/GTP-binding site motif A (P-loop) [AEAYLGKS].

Heat shock hsp70 proteins family signatures [IDLGTNS].

Hsp70\_2 [VFDLGGTFDVSI], Hsp70\_3 [VLLVGMTRVPKVOE].

contains EST gb:T4452, Z24523, T04782, Z17893, Z17973"

/codon\_start=1

/product="heat shock protein 70 like protein"

/protein\_id="CAB37531.1"

/db\_xref="GI:4467097"

/translation="MASVSAPKSVSANGKSMFGKLGVLARPCSPVGNVDVIGIDLG  
TTNSCVSMGKPTARVIEAEGSRTTPSVAMKQKCELLVGTAKRAQVNTNPTIFG  
SKRLIGRRFDDPOTQKEMKVPYKIYKAPNGDAWANGKFSQIGANVLTKMRET  
AEAYLGKSINKAVYTPAYFNDQAQRQATKAGTAGLDVQRI NEPTAAALSVMNNK  
EGVIAVFDLGGTFDVSI EISSGVFEVKATNGDTFLGDFDNTLLEYLVNFKKSD  
NIDLTDLNALQRLREAAEKATIELSTQTEINLPTITADACAKHLNLTLSKPKPE  
GLVKKLIERTRSPQNCCLKDAGVTIKEDVLLVGGTRVPKVOEIVSEIFGKSPCKG  
VNPDEAVANGAAIQGGILRGVDKDLLLDVPLSLGLETGLAVETKLI PENTTIPPKK  
SQVSTAAADNOVGKVLQGERMAADNKLGEFOLVGI PPAPRGMPQIEVTFDIDA  
NGITTVSKAKDQKTEQNTIRSSGSLSDDEINRMKAEALNAQKQKQLDLRNS  
ADPTIYVSKSLSEYREKIPAEIATSEIETAVSLDRTAMAGEDVEDIKAKVEAANKAVS  
KIEHMSKSGSGSGSGSGSGTGEQTPAEFEFGSGSRK"

complement(13386..14030)

/gene="F20D10.30"

/number=1

complement(14031..14117)

/gene="F20D10.30"

/number=1

complement(14118..14738)

/gene="F20D10.30"

/number=2

complement(14739..14827)

/gene="F20D10.30"

/number=2

complement(14828..15394)

/gene="F20D10.30"

/number=3

complement(15395..15490)



```
/gene="F20D10.30"
/number=3
complement(15491..15569)
/gene="F20D10.30"
/number=4
complement(15570..15668)
/gene="F20D10.30"
/number=4
complement(15669..15730)
/gene="F20D10.30"
/number=5
complement(15731..16042)
/gene="F20D10.30"
/number=5
complement(16043..16069)
/gene="F20D10.30"
/number=6
16556..19503
/gene="F20D10.40"
complement(16556..16790)
/gene="F20D10.40"
/number=1
complement(join(16556..16790,16869..17056,17133..17377,
17453..17618,17710..17975,18152..18358,18789..19503))
/gene="F20D10.40"
complement(join(16556..16790,16869..17056,17133..17377,
17453..17618,17710..17975,18152..18358,18789..19503))

alignment_scores:
  Quality: 1133.50      Length: 1065
  Ratio: 2.806          Gaps: 27
  Percent Similarity: 37.934  Percent Identity: 31.174

alignment_block:
US-09-591-466C-2 x ATF20D10/rev

Align seg 1/1 to reverse of: ATF20D10 from: 1 to: 147215

9 AspLeuArgTyrLeuLeuValValAlaLeuAlaPheIleTyrIle.. 24
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123209 GACTGTGAGATTCTCTCATCCGGCAGCTTTCATGTTCATCTACATCCA 123160
24 .....
123159 GGTCAATCTATATCCGATGATCGCTTTTGAAGTTGAAACTGAAATG 123110
24 .....
123109 ATTCATCGCTGCAACCTAATCTGTCTGTAAATATCCTCTCTCTTC 123060
25 Gln..MetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAla 40
::: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
123059 TCTAGATGAGGCTTTTCCAGACGAATACAGTATGAGATCGCTCAGT 123010
41 AlaAla.....
:::||||
123009 TCCGCTGTGAGTTCTTACCTTTTACCACCTGCGCTTTGTCTTTAGACCATG 122960
42 .....
122959 CTACTAGTTCCTTTTGAATCCCATTTTTTGGCTCTTACAGCCCATGA 122910
42 .....
122909 ACTACTTTCAAGAAGAGACACTATCTGCACATTATACACCATATCTTCG 122860
42 .....
122859 CTTTGCTATTTTCGATTTTTCCTGGTGTGTGATGAAGATCTATACACAGC 122810
43 .....
122809 CTTTCATTTGTTTCTCCAGATGAATCTGAGAACCATTGCACTAGTCAAA 122760
```

```
53 hrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgValValAla 69
||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
122759 TCGGAGGCTCTATAGTAGAGTTAGCATCAAAACAGTCGCGATTGTGCC 122710
70 LeuGlu..Glu.....
|||||
122709 CTCGAAGGTATATCCCTTCTCTTCTCATTTGGGACTTTGTGGAATCCT 122660
72 .....
122659 AGATGACCATGGTTCTTGGAACTTATTGAATACATAAAAAATTAACTTT 122610
73 .....GlnM 74
122609 CTAAATGGAGTAGAAGACTTATACATTGATTCTGAGTGTCTGTGTAGATA 122560
74 etLysHis...GlnAspGlnGluCysArgGlnLeuArgAlaLeuValGln 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122559 TGAAGAACCGCCAGGAGAACTTGTGCAGCTTAAGGATCTAATCCAG 122510
90 AspLeuGluSer.....
:::|||||
122509 ACGTTTGAAGTGCCTTACTATCATCCTGTAAACTCATTTCTTGCCTTC 122460
94 .....Lys.Gly 95
122459 TTTTATTAGTTTCTACTAATGCTTATTCTTCTCTCCAGAAAGGA 122410
96 IleLysLysLeuIle.....GlyAspVal.....
||| |||||||:|||||:|||||:|||||:|||||:|||||:
122409 ATAGAAACTCTACTCAAGGTGGACAGGCTGGATAAGAAGATGCCTTTC 122360
103 .....
122359 TTCGTATTGTAACTGCTATCATCTGTATATATATATACATCTAACCTGTT 122310
104 .....GlnMetProValAlaAlaValValValMetAla 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122309 TCTGTGTTTTTATTAGCAGATGCCTGTGCTGCTGTAGTGGTTATGCC 122260
115 CysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeu..... 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122259 TGCAGTGTGCACACTATCTTGAAGAGACTGTAAATACAGTTTAAAGTG 122210
129 .....
122209 AGATACTATACTGATTAAATCTTATACCACATTTTGTGAACTATCTA 122160
130 .....Lys..TyrGlnThrSerValAlaSerLysT 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122159 ATGGCTATTTCTTTTATACAGATATCAAACTCCCGTTGTCTCAAAAT 122110
139 ytrProLeuPheIle.....
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122109 ATCCTCTATTATATCTCAGGTACTTAATTTCCAGGCTGGTGGTCTCTTA 122060
143 .....
122059 GACTGTTTTATATCAACGAATTACTAACCATGGTCTTTTCCAACGACAT 122010
144 ....Ser..GlnAspGlySerAsnProaspValArgLysLeuAlaLeuSe 158
||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
122009 ATTTTCTGACAGGATGGATCTGATCAAGCTGTCAAGAGCAGCATCATGAG 121960
158 rTyrGlyGlnLeuThrTyrMet.Gln.....
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121959 CTATAATCAATTAACATATATGTCAGGTAATTTAATCTAGCGGTGATTTTC 121910
166 .....
121909 TCCTCACTGTTCAAGTAGATAAAATGAGGATGAAGGCTCGTTTTCCCTTT 121860
```

167 ....HisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182  
|||||:::||||||| |||||:::|||||  
121859 TCAGCACTCGATTTTGAACAGTGCTACTGAAAGGCCTGGCRACTGA 121810  
  
182 alAlaTyrrLysIleAlaAq.  
:::|||||:::|||||  
121809 CTGGCTACTACAAGATGACAGTAAGAGCTGGTTCTCTTTTCTCTGGCT 121760  
  
190 .....HisTyrrLysr 193  
121759 CCCTCCAGCTCCATTACTCAAACACTTTTTTAATACAGGTCACTACAAGTG 121710  
  
193 pAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIlelleL 210  
|||||:::|||||:::|||||:::|||||:::|||||  
121709 GGCACCTGGACCAGTTGTTTTCACAAAACAAATTTAGTCAGTGATTATAC 121660  
  
210 euGlu..... 211  
|||||  
121659 TAGAAGGTGGCGCTTTTAACTTTTCACACTCATCGTACCTTGGCTGTGT 121610  
  
211 ..... 211  
121609 ATTGTTTAAGAGGGCGTGTCTTTTACCCTTGTGGAGTATTGCTTAAGAGAGG 121560  
  
212 ..... AspAspMetGluIleAlaLaaspPheA 222  
|||||:::|||||:::|||||:::|||||  
121559 TTATCTGTCTCTTTAACAGATGATATGGAATTTGCTCCAAGACTCTTTG 121510  
  
222 spTyrrPheGluAlaGlyAlaThrLeuLeuAspArgAspLys..... 235  
|||||:::|||||:::|||||:::|||||  
121509 ATTACTTTGAGGCTCGAGCTAGTCTCATGGATAGGGATAAGTAGTTTACC 121460  
  
235 ..... 235  
121459 TACTGAAGCTAGACTCTTATAGTATAGTTTCCACAGAGTATACAAATGT 121410  
  
236 .....SerI 237  
:::  
121409 TTTTATTACTGTTTCTCATTCTTTTCTGGTCCAAACATGTTTAGAACCA 121360  
  
237 leMeAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGln 253  
|||||:::|||||:::|||||:::|||||:::|||||  
121359 TTATGGCTGCTTCATCATGGAATGATTAATGGACAGAAGCAGTTTGTGCAT 121310  
  
254 AspPro..... 255  
|||||  
121309 GATCCCTGTAAAGATTCTTCTTCTTCTTCTTACCTTGCTTCTCTCT 121260  
  
255 ..... 255  
121259 TGCTGTGTTCTTCATACCTCATATTATGCTAGTCTATGCCAGCTAAAAATA 121210  
  
256 ..... Asp..AlaLeuTyrrArgSerAspPheProGlyLeuGlyTrpM 270  
::: |||||:::|||||:::|||||:::|||||  
121209 ATCTGCAGATGCGCTATACCGATCAGATTTTCTCGCCTGGGTGGA 121160  
  
270 etLeuSerLysSerThrTrpSerGlnLeuSerProLysrPrpProlys... 285  
|||||:::|||||:::|||||:::|||||  
121159 TGCTCAAGAGATCGACTTGGATGAGTTATCACCAAGTGGCCAAAGCA 121110  
  
285 ..... 285  
121109 TATCCTTATATGCTTTTTTTTCTTTCAATCTCGATCACTATAGTTTGAGG 121060  
  
285 ..... 285  
121059 CTTCTTAAATCTAAAAAATAAATAGTTTGACGTTTCTTAAGACTTTTCT 121010  
  
286 ..... Ala.TyrrTrpAspTrpLeuArgLeuLysGluAsnHi 298  
|||||:::|||||:::|||||:::|||||  
121009 TAACAAATATCCACTTACTGGGATGATTGGCTGACACTAAAGGAAAAACA 120960  
  
298 sArgGlyArgGlnPheIleArgProGluValCysArgThrTyrrAsnPheG 315

I:::|||||  
120959 TAAAGCGCCAATTCATTGACCGGAAGTCTGTGAACAATCAAAATTTG 120910  
315 yGluHis..... 317  
120909 GTGAACATGTTAGTACCAGATTATGCTATCCTATATATCATAGTTTTACA 120860  
317 ..... 317  
120859 AGTCTGTTTTGTATGCTGTATAAATTGACCTATGCTGCATCATGAC 120810  
317 ..... 317  
120809 TTCTGCTTTCTGTATGAAAATGTTATATGACGCTCGTCTATAGTTT 120760  
317 ..... 317  
120759 CCCGACACATGGATGTAGACAGTCAATACCACATAATATATTTTGCATT 120710  
318 ..... GlycerSerLe 321  
120709 TCATACTTATCCATAAACCTGTGTATTAATTTGGCAGGGTCT.AGTTT 120661  
321 uGlyGlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspVal. 337  
120660 GGCACAGTTTTTTCAGTCTAGTATCTGGAACCTATAAAGCTAAACCATGTGA 120611  
337 ..... 337  
120610 CGGTAAGCGTAAGTGTTGTTTTTACCCTCTTTGTTGCATCACTAGTTT 120561  
338 ..... GlnValAspTrpLysSerMetAspLe 346  
120560 CACTGGTTTCCATTTCATTCATAAAGGTTGACTGGAAAGCAAGGACCT 120511  
346 uSerTyrLeuLeuGlu..... 351  
120510 GGGATACCTGCACAGAGGTAACATTTACAGAAGCTTCGGCCCTTTTTTTTA 120461  
351 ..... 351  
120460 TACCATTTTGCATGTTTWTGAGTGGAGAAATGTAGTTTTTAACCTGTCTC 120411  
352 ..... AspAsnTyrValLysHisPheGlyAspLeuVally 363  
120410 TCATGCTCTATAAAGGAAACTATACCAAGTACTTTTCTGGCTTAGTGAG 120361  
363 sLYSAIalysproIleHisGlyAlaaspalaValLeuLysAlaPheAsnI 380  
120360 ACAAGCACGACCAATTCAGAGTTCTGACCTTGCTTTAAAGGCTCAAACA 120311  
380 leAspGlyAspValArgIleGlnTyrArgAspGlnLeuAspPheGluAsp 396  
120310 TAAAGTAGATGTTGCTATCCGCTATAAGACCAAGTAGAGTTTGACGC 120261  
397 IleAlaArgGlnPheGlyIlePheGluGluTrpLys..... 408  
120260 ATTGCAGGGAATTTGCTATATTGAAGAACTGGAAGTAAATTTGAGATG 120211  
408 ..... 408  
120210 ATTGAACTATAATCTTCAATAATTGAGTCCAAATGGGTCACTTTTTT 120161  
409 ..... AspGlyValProArgAlaIalyLysGly 418  
120160 GTAACCTTTGATTGTGTAGGATGTGTGCTCGAACAGCATATAAGGA 120111  
419 IleValValPheArgPheGlnThrSerArgArgValPheLeuValSerPr 435  
120110 GTAGTGGTGTTCGATCCAGACACAGAGCTGTATTCCTGGTGGGCC 120061  
435 oAspSerLeuArgGlnLeuGlyValGluAspThr 446  
|||||||

```

120060 AGATTCTGTAATCCACCTTGGAAATTCGAAATTC 120027
seq_name: gb_pl:ATCHRIV89

seq_documentation_block:
LOCUS   ATCHRIV89
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 89.
ACCESSION AL161593
VERSION   AL161593.2 GI:7270796
KEYWORDS
SOURCE    thale cress.
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS  Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
JOURNAL  Unpublished
REFERENCE
AUTHORS  Robben,J., Braeken,M., Gymonprez,B., Volckaert,G., Mewes,H.W.,
Lemcke,K., and Mayer,K.F.X.
JOURNAL  Unpublished
REFERENCE
AUTHORS  Wedler,H., Kutzner,M., Wambutt,R., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
JOURNAL  Unpublished
REFERENCE
AUTHORS  4 (bases 1 to 199789)
EU Arabidopsis sequencing,project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV88 at the 5' end and an
overlap with ATCHRIV90 at the 3' end.

FEATURES             source
     source
exon                  1..199789
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     complement(5433..5537)
                     /gene="AT4g38130"
                     /number=1
                     5433..7997
gene                  /gene="AT4g38130"
                     complement(join(5433..5537,5612..5683,6112..6169,
                     6258..6328,6584..6814,6945..7385,7470..7997))
                     /gene="AT4g38130"
                     complement(join(5433..5537,5612..5683,6112..6169,
                     6258..6328,6584..6814,6945..7385,7470..7997))
                     /gene="AT4g38130"
                     /note="contains EST gb:R90034, AI994124.1, W43397,
                     AW004334.1, F15302, AA605428"
                     /codon_start=1
                     /product="Histone deacetylase"
                     /protein_id="CAB80478.1"
                     /db_xref="GI:7270797"
                     /translation="MDTGGNSLGGSDGVKRVKVCYFYDPEVGNVYVYGGHPMKPHRIR
                     MTHALLHYGLQHMVLKPPARDRLCRFHADYVFLRSITPEQQOQIRQLKRF
                     NVGDCPVFDGLYSCFOYTAGSVGGSVKLNHGLCDIAINWAGLHHKCEASGFCY
                     VNDIVLAILLELLKHQHVRLVDIDIHGGDVEEAFYATDRVMTVPFHKFGDYFPGTGH
                     IQDIGSGKTYLSNVPDLDGIDDESVEHLLFKPIMGKVMETFRPGAVVLOCGADSLSG
                     KMPGHEYIEYFGPDYTLHVPASNNKNSRQMLEEIRNDLLHNLKLOHAPSVPFOER
                     PPDTPEVDEDDGDKRWDPSMDVDDDRKPISRSVKREAVEPTDKDKGLKIM
                     ERGKGCEVEVDESGSTKVTGVPNVPVVEASVKMEEESTNKGGAQFAFPPT"

     source
exon                  complement(5538..5611)
                     /gene="AT4g38130"
                     /number=1
gene                  complement(5612..5683)
                     /gene="AT4g38130"
                     /number=2
                     complement(5684..6111)
                     /gene="AT4g38130"
                     /number=2
                     complement(6112..6169)
                     /gene="AT4g38130"
                     /number=3
                     complement(6170..6257)
                     /gene="AT4g38130"
                     /number=3
                     complement(6258..6328)
                     /gene="AT4g38130"
                     /number=4
                     complement(6329..6583)
                     /gene="AT4g38130"
                     /number=4
                     complement(6584..6814)
                     /gene="AT4g38130"
                     /number=5
                     complement(6815..6944)
                     /gene="AT4g38130"
                     /number=5
                     complement(6945..7385)
                     /gene="AT4g38130"
                     /number=6
                     complement(7386..7469)
                     /gene="AT4g38130"
                     /number=6
                     complement(7470..7997)
                     /gene="AT4g38130"
                     /number=7
                     8008..9245
                     /gene="AT4g38140"
                     complement(8808..9245)
                     /gene="AT4g38140"
                     /number=1
                     complement(8808..9245)
                     /gene="AT4g38140"
                     /note="similarity to RING-H2 zinc finger protein RHALA
                     -Arabidopsis thaliana, PID:g3790554"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80479.1"
                     /db_xref="GI:7270798"
                     /translation="MISMLFPRSPCLCTAIVFYTCVPIGLRLKKGDDADAHDDG
                     NLVGMFGDKKEEICCICLVFEAEADAVTHLPRCAHLPHINCIEPWLRLGHLTCP
                     LCRSFVLAPTPPTQNVNNAHSSSTLYLSIFFFCIFLHLGL"
                     /gene="AT4g38150"
                     10151..11059
                     /gene="AT4g38150"
                     complement(10151..11059)
                     /gene="AT4g38150"
                     /number=1
                     complement(10151..11059)
                     /gene="AT4g38150"
                     /note="similarity to leaf protein (Pnc401) -Pharbitis nil,
                     PID:d1026846"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80480.1"
                     /db_xref="GI:7270799"
                     /translation="MVPSSKAVVFARQMAKQIRVTPPSMSTRFLSTGDNQGVDEQON
                     PPEPLNPLRGRSSNSHREPPARQAHNLGKSDTLLSDGFLQFQKLVNQDSRETP
                     KPEQYFPEPLPPEDSEIFKKMKEGGLIPNAVMDLGLCKDGLVQEAQKFLGLMRDK
                     GTIPEVVIYVAVVEAFCKAHKIEDAKRIFRKNQNGIAPNFAFSYGLVQGLYNCLMD

```



115 CysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeu..... 129  
 43181 TGCAGTCGTCAGACTATCTTGAAGAGGACTGTAAATCAGTTTAAACGTG 43132  
 129 ..... 129  
 43131 AGATACTATCTGATTAAATCTTATACACAAATTTTGATTGAACATATCTA 43082  
 130 .....Lys..TyrGlnThrSerValAlaSerLysT 139  
 43081 ATGGCTATTCTTTTTTATACAGATATCAAACCTCCCGTTCCTCAAAT 43032  
 139 yrProLeuPheIle..... 143  
 43031 ATCCTCTATTATATCTCAGGTACTTAATTTGCAGGGTGGTGTCTCTTA 42982  
 143 ..... 143  
 42981 GACTGTTTTATATCAACGAATTAACCATGGTCTTTTCCAAGGACAT 42932  
 144 ....Ser..GlnAspGlySerAsnProaspValArgLysLeuAlaLeuSe 158  
 42931 ATTTCTCAGAGATGGATCTGATCAAGCTGTCAAGAGCAAGTCATTGAG 42882  
 158 rTyrGlyGlnLeuThrTyrMet.Gln..... 166  
 42881 CTATATCAATTAACATATATGCAGGTAATTTAATCTAGCGTGATTTTC 42832  
 166 ..... 166  
 42831 TCCTCACTGTTCAAGTAGATAAATGAGGATCAAGGCTCGTTTTCCCTTT 42782  
 167 ...HisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182  
 42781 TCAGCACTTGGATTTGAACAGGTGGTCACCTGAAAGGCGCTGGCAACTGA 42732  
 182 aIalaTyrTyrLysIlealaArg..... 189  
 42731 CTGCGCTACAAAGATTCGAGTAAGGACTGGTTCTCTTTCTCTCGCT 42682  
 190 .....HisTyrLysTr 193  
 42681 CCTCCAGCTCCTACTTGAACACTTTTAAATACAGGTCACTACAAGT 42632  
 193 pAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleL 210  
 42631 GGCACGTGGACCAAGTTGTTTTACAACACAAATTTAGTCGACTGATTATAC 42582  
 210 euGlu..... 211  
 42581 TAGAAGTGGCGCTTTTAACTTTACACTTCACACTCAGTACCTTGTGTGT 42532  
 211 ..... 211  
 42531 ATTGTTTAAGAGGGCTGCTTTTACCCTTGTGGAGTATTGCTTAAGAGG 42482  
 212 .....AspAspMetGluIleAlaAlaAspPhePheA 222  
 42481 TTATCTCTGCTCTTTAAACAGATGATATGGAATTTGCTCCAGACTCTTTG 42432  
 222 sPTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLys..... 235  
 42431 ATTACTTTGAGCTGCAGCTAGTCTCATGGATAGGATAAGTATGTTACC 42382  
 235 ..... 235  
 42381 TACTGAAGCTAGACTCTTATAGTATAAGTTTCCACAGAGTATACAAATGT 42332  
 236 .....SerI 237  
 42331 TTTTATTACTGTTTCTCTCTTTCTGGTCCAAACATGTGTTAGAACCA 42282  
 237 leMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGln 253

42281	TTTTTTTTT TTATGCTGCTTCATCATGGAATGATAATGGACAGAAGCAGTTGTGCAT	..... 	42282	..... 
254	AspPro.....	..... 	255	..... 
42231	GATCCCTGTAAGATTCTTCTTCTTCTTCTTACCTTGCCTCCTCTCT	..... 	42182	..... 
255	.....	..... 	255	..... 
42181	TGCCTGTCTTCAATACCTCAATATATGCTAGCTCATCGCAGCTAAAAATA	..... 	42132	..... 
256	.....Asp...AlaLeuTyrArgSerAspPheProGlyLeuGlyTrpM	..... 	270	..... 
42131	ATCTGCAGATGCCTATACCGATCAGATTTTCTTCTGGCCTTGGTGGA	..... 	42082	..... 
270	eLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProLys...	..... 	285	..... 
42081	TGCTCAAGAGATCGACTTGGGATGAGTTATCACCAAAAGTGCCAAAGCCA	..... 	42032	..... 
285	.....	..... 	285	..... 
42031	TATCCTTATATGCTTTTTTTTCTTTCAATCTCGATGACTATAGTTGACG	..... 	41982	..... 
285	.....	..... 	285	..... 
41981	CTTCTTAAATTTCTAAAAAATAATAGTTTGACGTTTCTTAAGACTTTTCT	..... 	41932	..... 
286	.....Ala...TyrTrpAspAspTrpLeuArgLeuLysGluAsnHi	..... 	298	..... 
41931	TAACAATAATCCACTTACTGGGATGATTGGCTGAGACTTAAGGAAACCA	..... 	41882	..... 
298	sArgGlyArgGlnPheIleArgProGluValCysArgThrTyrAsnPheG	..... 	315	..... 
41881	TAAAGGCGCGCAATTCATTCACCGGAAGTGTGAGAACATACAATTTTG	..... 	41832	..... 
315	lyGluHis.....	..... 	317	..... 
41831	GTGACATGTTAGTACCCGATTATGCTATCCTATATATCATAGTTTACA	..... 	41782	..... 
317	.....	..... 	317	..... 
41781	AGTCTGTTTTTGTAAAGCTGTTAAATTGACCTATGTTGCTGCATATGAC	..... 	41732	..... 
317	.....	..... 	317	..... 
41731	TTCCTCTTCTGTATGAAAAATGATTATTGACAGCTTCGCCTATAGTTT	..... 	41682	..... 
317	.....	..... 	317	..... 
41681	CCCGACACATGGATGTAGACAGTCAATACACATAATATATTTTGCATT	..... 	41632	..... 
318	.....GlySerSerLe	..... 	321	..... 
41631	TCATACITATCCATAACCTGTGTATATTAATTTGGCAGGGGTCT.AGTTT	..... 	41583	..... 
321	uGlyGlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspVal.	..... 	337	..... 
41582	GGGACAGTTTTTCAGTCAGTATCTGGAACCTATAAAGCTAAACGATGTGA	..... 	41533	..... 
337	.....	..... 	337	..... 
41532	CGGTAACGGTAAGTGTGTTTTTTTACCCTTTTGTGTGCATCACTAGTTTT	..... 	41483	..... 
338	.....GlnValAspTrpLysSerMetAspLe	..... 	346	..... 
41482	CACTGGTTTCTTATTTTCATTAAGAGTTGACTGGAAAGCAAGCACCT	..... 	41433	..... 
346	uSerTyrLeuLeuGlu.....	..... 	351	..... 
41432	GGGATACCTCACAGAGGTAAACATTTACAGAACTCCGGCCTTTTTTTTTTA	..... 	41383	..... 
351	.....	..... 	351	..... 

```

41382 TACCAATTTTGCATGTTTTCAGTGGAGAAATGTAGTATTTTAAACCTGTTCTC 41333
352 .....AspAsnTyrVallyHisPheGlyAspLeuVally 363
41332 TCATGCTCATCAAGGAACTATACCAAGTACTTTTCGCTTAGTGAG 41283
363 slyAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnI 380
41282 ACAAGCAGCAACAATCAAGGTTCTGACCTTGCTTAAAGGCTCAAAACA 41233
380 leAspGlyAspValargIleGlnTyrArgAspGlnLeuAspPheGluAsp 396
41232 TAAAGGATCATGTTCTGATCCGATATAAGACCAAGTAGAGTTTGAACGC 41183
397 IleAlaArgGlnPheGlyIlePheGluGlnTrpLys..... 408
41182 ATTGCAGGGAATTTGGTATATTTGAGAATGAAGGTAATTTGAGATG 41133
408 ..... 408
41132 ATTGAACATAATCTTCAATAATTTAGTCCAAATGGTCACATTTT 41083
409 .....AspGlyValProArgAlaAlaTyrLysGly 418
41082 GTACCCCTTTGATTGTGTAGATGGTGTCCCTCGAACGACATATAAGGA 41033
419 IleValValPheArgPheGlnThrSerArgArgValPheLeuValSerPr 435
41032 GTAGTGGTGTTCGAATCCACACAAAGAGCGTGTATCTCTGGTGGGCC 40983
435 oAspSerLeuArgGlnLeuGlyValGluAspThr 446
40982 AGATTCTGTAATGACGCTTGAATTCGAAATTC 40949
seq_name: gb_pl:ATF22113

seq_documentation_block:
LOCUS ATF22113 93760 bp DNA linear PLN 27-AUG-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F22113 (ESSA
project).
ACCESSION AL035539
VERSION AL035539.1 GI:4539331
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 93760)
AUTHORS Bevan,M., Wedler,H., Kutzner,M., Wambutt,R., Bancroft,I.,
Mewes,H.W., Mayer,K.F.X. and Schueller,C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 93760)
AUTHORS EU Arabidopsis sequencing,project.
JOURNAL Direct Submission
TITLE Submitted (26-FEB-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schueller@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosome 4 can be
viewed at: http://webserv.mips.biochem.mpg.de/proj/thal/.
FEATURES
source
1..93760
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(2..106)
/gene="F22113.10"
/number=1
exon
complement(1858..1900)

```

```

/ gene="F22113.10"
/ number=8
complement(1901..1963)
/ gene="F22113.10"
/ number=9
complement(1964..2055)
/ gene="F22113.10"
/ number=9
complement(2056..2101)
/ gene="F22113.10"
/ number=10
complement(2102..2178)
/ gene="F22113.10"
/ number=10
complement(2179..2255)
/ gene="F22113.10"
/ number=11
complement(2256..2448)
/ gene="F22113.10"
/ number=11
complement(2449..2528)
/ gene="F22113.10"
/ number=12
complement(2529..2668)
/ gene="F22113.10"
/ number=12
complement(2669..2756)
/ gene="F22113.10"
/ number=13
complement(2757..2969)
/ gene="F22113.10"
/ number=13
complement(2970..3020)
/ gene="F22113.10"
/ number=14
complement(3021..3124)
/ gene="F22113.10"
/ number=14
complement(3125..3193)
/ gene="F22113.10"
/ number=15
3517..4827
/ gene="F22113.20"
/ number=1
3517..4827
/ gene="F22113.20"
3517..4827
/ gene="F22113.20"
/ note="similarity to amino acid transport protein -
Arabidopsis thaliana, PID:g2576363
contains EST gb:T43079, T20761, T41914, Aa586275, T43892"
/ codon_start=1
/ product="putative amino acid transport protein"
/ protein_id="CAB37481.1"
/ db_xref="GI:4539333"
/ translation="MGONEASSSSYTLKIPPPARESPLLGKGPPLSSQPKTFANVF
IAYGAGVLGPLYAFKRTGWLGMVLLVSVSLVTHHGMLLVYTRRKLDGDFNAGISKI
GSRDGLFAGVCGSLGRIVVDLFIILSQAGFCVGLIFIGTTLANLSDPESPTSLRHQF
TRLGSEFGVSSKSLIWCFCFQGLNSKLTFLHAPLSIFADIVDLGAMAVVIED
SMILKQPDVAVFGMSLFYFGMVAVYFEGVGMVLPLESEMKKDKFKGVLLGCM
GFIILYTAFIGILGYLAFGEDTMDIITANLGLAVSVVOLGLCINLFTFPLMNPV
FEIVRRFRGMYSAWLRLVLAIVLAVLVPNFADFLSVGSSTCCVLGLFVLPALE
HLVFKEMGHLWSSDTAIVLGVLAIVSGTSSSEIFSVKV"
complement(5061..5222)
/ gene="F22113.30"
/ number=1
complement(join(5061..5222,5303..5526,5567..5777,
5859..6034,6112..6127))
/ gene="F22113.30"
5061..6127
/ gene="F22113.30"
complement(join(5061..5222,5303..5526,5567..5777,
5859..6034,6112..6127))

/ gene="F22113.30"
/ note="similarity to gene T10 protein - mouse,
PIR2:S37488"
/ codon_start=1
/ product="putative protein"
/ protein_id="CAB37482.1"
/ db_xref="GI:4539334"
/ translation="MKITTEALRWEDGETVGGGRDLVGGTWNLCGTRHGRLAFLTNFK
EASSFPAKSRGDLPLRYLQSEKSPAEAEETQDEISLYGNLVLVYHVLKSMIYIT
NRPFGDKLVTVQSPGIVHLSNANLSPKPNVSNPCVFLSLQCLRLREGFQOLLAEN
GSGEPFKTMVEEVNTVKTDETELPHTVETETVYHLSIFVDMORPTGRYGTSSIS
AIVKSHGSDGGDGEICFYERHLEEGDSNKEHTQQFVLIQNQSI"
complement(5223..5302)
/ gene="F22113.30"
/ number=1
complement(5303..5526)
/ gene="F22113.30"
/ number=2

alignment_scores:
Quality: 1131.50 Length: 1062
Ratio: 2.822 Gaps: 27
Percent Similarity: 37.759 Percent Identity: 31.262

alignment_block:
US-09-591-466c-2 x ATF22113/rev ..
Align seg 1/1 to reverse of: ATF22113 from: 1 to: 93760
9 AsLeuArgTyrLeuLeuValAlaAlaLeuAlaPheIleTyrIle.. 24
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3175 GACTTGAGATTCTTCTCATCCCGCAGCTTTCATGTTTCATACATCCA 3126
24 ..... 24
3125 GGTCATCTATATCCGATGATCTCGCTTTTGAAGTTGAACCTGAATTC 3076
24 ..... 24
3075 ATTCCATCGCTGCAACCTAAATCTGTCTGTAATTATCTCTCTCTCTTC 3026
25 Gln..MetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAla 40
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
3025 TCTAGATGAGGCTTTCCAGACCAATCACAGTATGCAGATCGCCTCAGT 2976
41 AlaAla..... 42
::|||
2975 TCCGCTGTGAGTTCTTACTTTTACCACCTGCCCTTTTGTCTTTAGACCATTC 2926
42 ..... 42
2925 CTACTAGTTCCTTTTGAATCCCAATTTTGGCTCTTTACAGCCCATGA 2876
42 ..... 42
2875 ACTACTTTCAAGAAGAAGACACTATCTGCACATTACACCATATCATTCG 2826
42 ..... 42
2825 CTTTGGTATTTCGATTTTTCCTGGTGTGTGATGAAGATCTATACACAGC 2776
43 ..... 53
|||||:|||||:|||||:|||||:|||||:|||||:
2775 CTTTCATTGTGTTCTCCAGATCGAATCTGGAACCATTCGACATAGTCAAA 2726
53 hrArgLeuLeuIleAspIysIleSerGlnGlnGlnIleValValAla 69
||| |||||:|||||:|||||:|||||:|||||:|||||:
2725 TCCGAGGCCTCATAGATGAAGTTAGCATCAAAACAGTCGCGGATGTTGCC 2676
70 LeuGlu..Glu..... 72
|||||
2675 CTCGAAGGTATATCCCTCTCTTCTTCTCCTCATTTGGGACTTGTGGAATCCT 2626

```

```

72 ..... 72
2625 AGATGACCATGTTCTTGGAACTATTGAATACATATACAAAATAACTTT 2576
73 .....GlnM 74
2575 CTAATGGAGTAGAAGACTTATACATTGATTCTGAGTGTCTTGATAGA 2526
74 etLysHis...GlnAspGlnGluCysArgGlnLeuArgAlaLeuValGln 89
2525 TGAAGAACCGCCAGGACGACGAAGAACTTGTGCAGCTTAAGGATCTAATCCAG 2476
90 AspLeuGluSer..... 93
2475 ACGTTTGAAGTGGCTTACTATACCTGTAAAACTCATTTTCTTGGCTTC 2426
94 .....Lys.Gly 95
2425 TTTTATTAGTTTTCTACTAATGCTTATTCTTCTCCAGAAAAAGGA 2376
96 IleLysLysLeuIle.....GlyAspVal..... 103
2375 ATAGCAAAACTCACTCAAGGTGGACAGTCTGGATAAGAAGATGCCTTTC 2326
103 ..... 103
2325 TTCGTATTGTAACTGCTACTTGTCTATATATATACATCTAACCTGTT 2276
104 .....GlnMetProValAlaAlaValValMetAla 114
2275 TCTTGTTTTTATTAGCAGATCCCTGTGGCTGCTGTGAGTGGTTATGGCC 2226
115 CysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeu.... 129
2225 TGCAGTGTGCAGACTATCTTGAAGGACTGTAAATCAGTTTTAACGTG 2176
129 ..... 129
2175 AGATACTACTGATTAAATCTTATACCACATTTTGATTGAACATATCTA 2126
130 .....Lys..TyrGlnThrSerValAlaSerLysT 139
2125 ATGGCTATTCTTTTTTATACAGATATCAAACTCCGTTGCTTCAAAAT 2076
139 yrProLeuPheIle..... 143
2075 AUCCTCTATTATATCTCAGTACTTAATTTGCAGGGTGGTGGTCTCTTA 2026
143 ..... 143
2025 GACTGTTTTATATCAACGAATTACTAACCATGGTCTTTTCCAAAGCAT 1976
144 ....Ser..GlnAspGlySerAsnProAspValArgLysLeuAlaLeuSe 158
1975 ATTTCTGACAGGATGGATCTGTCAAGGTGTCAAGAGCAAGTCATTGAG 1926
158 rTyrGlyGlnLeuThrTyrMet.Gln..... 166
1925 CTATAATCAATTAACATATATGCAAGTAATTAATCTAGCGTGAATTTTC 1876
166 ..... 166
1875 TCCCTCACTGTTCAAGTAGATAAATGAGGATGAAGCTCGTTTCCCTCTT 1826
167 ....HisLeuAspTyrGluProValHisThrGluArgProGlyGluLeu 182
1825 TCAGCATTGATTTTGAACAGTGGTGCATCTGAAAGGCCCTGGCGAATGA 1776
182 alaIaTyrTyrLysIleAlaArg..... 189
1775 CTGCGTACTACAAGATGCAGGTAAAGGACTGGTTTCTCTTTCTCTCGCGT 1726
190 .....HisTyrLysTr 193
1725 CCCTCCAGCTCCATTACTGAACACTTTTAAATACAGGTCACACAAGTG 1676
193 pAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleL 210
1675 GGCACCTGGACAGTTGTTTTACAAACACAAATTTAGTCGAGTATTATAC 1626
210 euGlu..... 211
1625 TAGAAGGTGGCGCTTTTAAACCTTTTCACACTCATCGTACCTTGTGTGT 1576
211 ..... 211
1575 ATTGTTTAAGAGGGGCTGTCTTTTACCTTGTGGAGTATTGCTTAAGAGAG 1526
212 .....AspAspMetGluIleAlaAlaAspPhePheA 222
1525 TTATCTTGCTCTTTAACAGATGATATGGAATTTGCTCCAGAGCTTCTTTG 1476
222 spTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLys..... 235
1475 ATTACTTTGAGGCTGCAGCTAGTCTCATGTAGGATAGTATGTTACC 1426
235 ..... 235
1425 TACTGAAGCTAGACTCTTATAGTATAAAGTTTCCACAGAGTATACAAATGT 1376
236 .....SerI 237
1375 TTTTATTACTGTTTCTCATCTTTTCTGGTCCAAACATGTGTTAGAACCA 1326
237 leMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGln 253
1325 TTATGGCTGCTTCATCATCGAATGATAATGGACAGAAGCAGTTTCTGCAT 1276
254 AspPro..... 255
1275 GATCCCTGTAAGATTCTTCTTCTTCTTCTTCTTACCTTGTCTCTCTCT 1226
255 ..... 255
1225 TGCTGTGTTTTCATACCTCATATTATGCTAGCTCATCGAGCTAAAATA 1176
256 .....Asp..AlaLeuTyrArgSerAspPheProGlyLeuGlyTrpM 270
1175 ATCTGCAGATCGCTATACCGATCAGATTTTTTCTCGCCCTTGGTGGA 1126
270 etLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProLys.... 285
1125 TGCTCAAGAGATCGACTGGGATGAGTTATCACCACAAAGTGGCCAAAGGCA 1076
285 ..... 285
1075 TATCCTTATATGCTTTTTTTTCTTCAATCTCGATGACTATAGTTTGACG 1026
285 ..... 285
1025 CTCTCTTAATTTCTAAAAAATAAATACTTTTGACGTTTCTTAAGACTTTTCT 976
286 .....Ala..TyrTrpAspAspTrpLeuArgLeuLysGluAsnHi 298
975 TAACAAATATCCACTTACTGGGATGATTGGCTCAGACTAAAGGAAACCA 926
298 sArgGlyArgGlnPheIleArgProGluValCysArgThrTyrAsnPhag 315
925 TAAAGGCCGCCAATTCATTTCGACCGGAAGTCTGTAGAACATACAATTTTG 876
315 lyGluHis..... 317
875 GTGAACATGTTAGTACCCGATTTATGCTATCTCTATATATCATAGTTTACA 826
317 ..... 317

```



```
825 AGTCTGTTTTTGTAAATGCTTTAAATGACCTATGTTCCGTCATATGAC 776
317 ..... 317
775 TTCTGCTTTCTGTATGAAAAATGATATTATTGCAGCTTCGTCCTATAGTTT 726
317 ..... 317
725 CCGCACATGGATGTAGACAGTCAATACACATAATATATATTTTGCATT 676
318 .....GlySerSerLe 321
675 TCATACTTATCCATAACCTTGTTATATTAATTTGGCAGGGTCT.AGTTT 627
321 uGlyGlnPhePheLysGlnTyrluGluProIleLysLeuAsnAspVal. 337
|||||..... 337
626 GGGACAGTTTTTCAGTCAGTATCTGGAACTATAAAAGCTAAACGATGTGA 577
337 ..... 337
576 CGGTAAGCGTAAGTGTGTTTTTTTACCCCTTTTGTGTCATCACTAGTTT 527
338 .....GlnValAspTrpLysSerMetAspLe 346
526 CACTGTTTTCCTATTCATTCATAAAGGTGACTGGAAAGCAAGACCT 477
346 uSerTyrluLeuGlu..... 351
476 GGGATACCTGACAGAGGTAACATTTTACAGAACTCCGGCCTTTTTTTT 427
351 ..... 351
426 TACCATTGTCATGTTTTCAGTGGAGAAATGAGTTTTTAACTGTTCTC 377
352 .....AspAsnTyrluValLysHisPheGlyAspLeuVally 363
376 TCATGCTCTATCAAGGGAACCTATACCAAGTACTTTCTGGCTTAGTGAG 327
363 sLysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnI 380
|||||..... 380
326 ACAAGCACGCCAAATCAAGGTTCTGACCTGTCTTAAGGCTCAAAACA 277
380 leAspGlyAspValArgIleGlnTyrluArgAspGlnLeuAspPheGluasp 396
|| ..... 396
276 TAAAGGATCATGTTTCGTATCCGGTATAAGACCAAGTAGAGTTTGAAGC 227
397 lleAlaArgGlnPheGlyllePheGluGluTrpLys..... 408
||||| ..... 408
226 ATTGCAGGGGAATTTGGTATATTTGAAGAATGGAAGGTAATTTGAGATG 177
408 ..... 408
176 ATTGAACATATAATCTTCATAATTTAGTTCGCAATGGTCACTTTTTT 127
409 .....AspGlyValProArgAlaAlaTyrluLysGly 418
|||||..... 418
126 GTACCCCTTTGATTGTGTAGGATGTTGTCGCTCGAAGACATATAAGA 77
419 lleValValPheArgPheGlnThrSerArgArgValPheLeuValSerPr 435
76 GTAGTGGTGTTCGAATCCAGACAACAGACGTGTATTCCTGGTTGGGCC 27
435 oAspSerLeuArgGlnLeuGlyVal 443
26 AGATTCTGTAATGCAGCTTGAATT 2
```

seq\_name: gb\_htg.AP004082

seq\_documentation\_block:

LOCUS AP004082 120538 bp DNA linear HTG 23-AUG-2001  
DEFINITION Oryza sativa chromosome 2 clone OJ1149\_C12, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AP004082

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AP004082.1 GI:15281361  
HTG; HTGS\_PHASE2.  
Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1149\_C12.  
Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 120538)  
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC  
clone:OJ1149\_C12

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Published Only in Database (2001) In press  
2 (bases 1 to 120538)  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Direct Submission  
Submitted (22-AUG-2001) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
The nucleotide sequence of this BAC clone was generated by  
combining Monsanto and RGP-Japan sequencing data.  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: this is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

COMMENT

Location/Qualifiers  
1..120538  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="2"  
/clone="OJ1149\_C12"  
BASE COUNT 33869 a 26177 c 26373 g 34016 t 103 others  
ORIGIN

FEATURES

Source

alignment\_scores:  
Quality: 886.00 Length: 611  
Ratio: 3.488 Gaps: 12  
Percent Similarity: 41.571 Percent Identity: 35.025  
alignment\_block:  
US-09-591-466C-2 x AP004082/rev ..  
Align seg 1/1 to reverse of: AP004082 from: 1 to: 120538

165 MetGlnHisLeuAspTyrluProValHisThrGluAtgProGlyGluLe 181  
:::|||||.....  
103602 TTGCAGCATTTAGATCTTGAACCTGTGCGCACTGAAGGCCAGGAGATT 103553  
181 uValalaTyrluTyrluSilleAla..... 188  
|::|.....  
103552 GATAGCATATTACAAGATAGCAGTAGGTGCTCCTTGTATTGATGATAT 103503  
188 ..... 188  
103502 CCGTATTGAAGATAGTGATGATACCACTGATCCTTGTCTGATACTTTC 103453  
189 .....Arg..HisTyrluSripAlaLeuAspGlnLeuPheHisLysHis 202  
::: |||||.....  
103452 TTATTTTCAGAACATTTACAAATGGCCTTGGATGAGTATTTCATCAACAC 103403  
203 AsnPheSerArgValIleIleLeuGlu..... 211  
|||||.....  
103402 AACTTTGCGGAGTAATAATTTCTGGAAGGTTGTACACCTTGGATGCTCTT 103353

211 ..... 211  
103352 ATAACATATTTCTCTTTGATCTGCTCTTTCCATCCCATTTATGGTA 103303  
212 ..... AspAspMetGluIleAlaAlaAspPhePheAspTyr 223  
103302 ATAATTTGCTAATAGATGATATGGAGATCGCCCTGATTTTGGACTAC 103253  
224 PheGluAlaGlyAlaThrLeuLeuAspArgAspLys ..... 235  
103252 TTCGAGGCTGCAGCTAAATTTACTTGATAACGATAAGTATGTTGATATTA 103203  
235 ..... 235  
103202 TATATTATATTAGTTGGAACCCCAAGGATTTCTCTCCAATGATCTAA 103153  
236 ..... SerIleMetAla 239  
103152 TTGCTTTTGCTGCTATTCCTGTCACCTGTGCTATGAAGGACATCATGGC 103103  
240 IleSerSerTrpAspAspAsnGlyClnArgGlnPheValGlnAspPro.. 255  
103102 GTTTCCTCTGGAATGACATGGACAAAAGCAATTTGTTACGACCCCAAG 103053  
255 ..... 255  
103052 TTAGTACTCTGATAGCCTGTATAAAGTTCTGTTCTCCATTATAGGA 103003  
256 ..... AspAlaLeuTyrArgSe 261  
103002 AAGCTACTCATTTTCCTACATAAACATTTTGCAGAAGCTCTTTACCGGTC 102953  
261 rAspPhePheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerG 278  
102952 GGATTTCTTCTGCTGGACTGGATGGATGTTAACAAAGCCTACATGGATTG 102903  
278 luLeuSerProLysTrpProLysAla ..... 286  
102902 AGCTGTCACTCAAGTGGCCCAAGCATATCTTTACTGTGCTACTAATT 102853  
286 ..... 286  
102852 TTTATATTAGCATATACAAATATTTTCCATCTACCATATATTTCTTCCTTA 102803  
287 ..... TyrTrpAspTrpLeuArgLeuLysGluAsnHis 299  
102802 ACAAGACTTTTACTTATTTGGATGACTGGTGAGGCTAAAGAGGTACACA 102753  
299 rgGlyArgGlnPheIleArgProGlyValCysArgThrTyrAsnPheGly 315  
102752 GAGACCGCAATTTATTCGCCCGGAGTCTGCAGAACATACAATTTTGGC 102703  
316 GluHis ..... 317  
102702 GAACATGATATGTTGGCTACTTCATCAATGCAAAATTAACCACATTTCTGC 102653  
317 ..... 317  
102652 ACACGTAGCCACTGGATAGTAATTTAGTCTTTGAGCCAGTAAGTATGG 102603  
318 ..... GlySer 319  
102602 ATGACTTGAATTTACATAATAGATTTTCATACCTCTCTTTTCAGGGGTCA 102553  
320 SerLeuGlyClnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAs 336  
102552 AGCATGGGCAATTTTCACACAGTACTTTGGAACCAATCAAGCTAAATGA 102503  
336 pValGlnValAspTrpLysSerMet ..... 344  
102502 TGCCCATGTATGCAAACTTCCATCACCCTTTGTTTGTAACTGTGTC 102453  
345 ..... AspLeuSerTyrLeuLeu ..... 350

102452 AAATTCCTTTAGTCAGAGATTTACCATTTTACCTCTCTCCTTTGTGGC 102403  
350 ..... 350  
102402 TATAGATCAAGTGAATTTCTCAGGACCTGAGCTACCTCAAGGAGTAAGA 102353  
350 ..... 350  
102352 ATTTGTTTGGACTTTGAACACAGTAGTTTGCAGCAAAATCCCTAATATTAT 102303  
351 ..... GluAspAsnTyrValLysHisPheGlyAspLeuValLys 363  
102302 GTTGGCAACACAGGACAAGTTCTGATCCAATTTGGAAAAGACGTCGCT 102253  
364 LysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnI 380  
102252 AGTCCACCCCTCTCCATGGATCTGATGCCGCGTTGAAAGCCCAATAT 102203  
380 eAspGlyAspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspI 397  
102202 GGATGCGGATGTAAGGATCCAGTATAACGATCAGGAAGACTTTTGAGCGGA 102153  
397 leAlaArgGlnPheGlyIlePheGluTrpLys ..... 408  
102152 TAGCTGCTCAATTTGGAATATTTGAAGATGGAAGGTTTACTTCTTTTGAC 102103  
408 ..... 408  
102102 CCCATTTTATTCATAAATAATGCTCTTGTCAAGTAGGTCAAGTGACAA 102053  
408 ..... 408  
102052 TTTTGTGCTGTGATTTTTCAGTGGTGATTTGTTAGTCTAGTAGAAAATGT 102003  
408 ..... 408  
102002 TTCAGTTTGCAGTTTTCATCTAGATAGACATCAGCCCCGCGTAATGGTCTA 101953  
408 ..... 408  
101952 TAATTTTCTTTGGACTTCTCTGTTTTTTTAGTGTGTGATGATCATGTAT 101903  
409 ..... AspGlyValProArgAlaAlaTyrLysG 418  
101902 GTCTGTTTCTCTGTTTTCAGATGGCATTCACAGAACAGCTTATAAAG 101853  
418 lyIleValValPheArgPheGlnThrSer...ArgArgValPheLeuVal 433  
101852 GAGTAGTAGTCTTCCGGTACAAGAGTAGTCAAGACGAATATACCTCGTT 101803  
434 SerProAspSerLeuArgGlnLeuGlyVal 443  
101802 GGTCGGATTCCTCAGTCAGCTTAGGGTA 101773